Quantum Contextual Information Access and Retrieval

Emanuele Di Buccio, Giorgio Di Nunzio, Massimo Melucci, and Nicola Orio*

University of Padua, Department of Information Engineering

Abstract. This paper illustrates the research work being in progress at the University of Padua within the VII FP Marie Curie International Research Staff Exchange Scheme project QONTEXT, which shows a new vision of information access and retrieval based on Quantum Theory. In particular, this paper describes the basics of QT, the use of the formalism for relevance feedback, music retrieval and visual cluster-based data mining.

1 Introduction

Information needs are becoming even more broad and complex essentially due to the context of the search process (e.g. the system, user, language, word meaning, *medium*, interface and interaction methods). The research in Information Access and Retrieval (IAR) had led to various search engine models, such as vector space and probabilistic models. Unfortunately there has been no comprehensive investigation at the theoretical level for effectively integrating elements of context to create advanced search technologies. The key issue preventing such research is a lack of a unified theoretical framework to seamlessly integrate the dimensions of context into the search engine models and the evaluation protocols.

QONTEXT is a research project participated by seven partners from Europe, Canada, China and Australia¹ who share the belief that the dimensions of context can be naturally integrated into a generic and fundamental framework. To address the challenges of the dimensions of context in IAR, QONTEXT shows a new vision of the IAR paradigm based on Quantum Theory (QT) based on [1]. QT allows to measure relevance and context via vector subspace projection. At the University of Padua, the research will deal with multimedia objects (e.g., music) without being anchored to textual descriptors, to modeling relevance feedback as context evolution using unary operators and to designing clusterbased visualization methods.

The work done by the QT-based IAR research community so far [2] is encouranging because it has shown the experimental feasibility, the powerful modeling formalism, the adaptability to various media and tasks, and suggests that there is room for improvements.

 $^{^{\}star}\ firstname.lastname@dei.unipd.it$

¹ Principal Investigators: Sven Aerts (Belgium), Peter Bruza (Australia), Yuexian Hou (China), Joemon Jose (UK), Massimo Melucci (Italy), Jian-Yun Nie (Canada), Dawei Song (UK)

Notion	Classical	Quantum
Event space	Ω	Hilbert vector space \mathcal{H}
Random event	Set	Orthonormal basis $\{ B\rangle, \bar{B}\rangle\}$
Probability Measure	Set measure	State vector $ \varphi\rangle$

Table 1. The correspondence between classical probability and quantum probability



Fig. 1. The correspondence between classical probability and quantum probability

2 The Basics

In the classical probabilistic model, events (e.g., word occurrences, category memberships, relevance, location, task, genre) are represented as sets and the probability measure is based on a set measure, e.g., set cardinality. In contrast, in quantum probability, events are represented as orthonormal vectors and the probability measure is the trace of the product between a density matrix and the matrix representing an event as summarized in Table 1. The simple example in Figure 1 depicts that when vectors are used to implement both events and densities the probability in the vector space is the squared inner product between the vectors, that is, the squared size of the projection of $|A\rangle$ onto $|\varphi\rangle$.

3 Relevance Feedback

Diverse are the sources of evidence can be adopted, e.g. the behavior of the user when interacting with the results or judgments explicitly provided by the user for Relevance Feedback (RF) purpose. The combination of diverse sources of evidence is complex because of the feature heterogeneity (e.g. term frequency for the documents content and display-time for the user interaction behavior) and the lack of knowledge of the factors affecting relevance assessment. QT can cope with this complexity because it provides a uniform formalism to model the diverse factors. In RF, a document can be represented as a density matrix. The factors may be extracted from the training documents using local co-occurrence data of terms to obtain a term correlation matrix, thus applying SVD to the matrix and obtaining a vector basis. Thus, the probability of relevance can be computed by the trace function. Factors can also model user location or different behavioral patterns extracted from the behavior of the user when interacting with the results. The main issue is to obtain a vector basis starting from the evidence gathered from each source of evidence and the mechanism for the basis computation should be able to unveil the most meaningful factors from the data. The main advantage is, from the one hand, the uniform modeling of diverse sources of evidence and, on the other hand, the greater generality of quantum probability than that of classical probability.

Our aim within QONTEXT is to employ QT for modeling RF within a complex and heterogenous feature space.

4 Music Retrieval and Processing

Music Information Retrieval (MIR) is an emerging research area that focus on providing new access methodologies and interaction paradigms to very large music collections. The wide availability of portable music players, paired by the increasing amount of digital music available to the end user, makes music access particularly related to context: music can be used as a background at work and study, as a distinctive for social groups, as a way to promote personal's image (especially within young generations). Moreover, it is generally acknowledge that neither textual metadata nor content-based descriptors alone can completely describe the music content and the user information need.

For these reasons, the application of QT to music access and retrieval is particularly promising. On the one hand, it will allow to model in a unified framework different sources of evidence, related both to the social role of music – i.e., genre, usage, user provided tags – and to the pure acoustic features – i.e., melody, rhythm, timbre. Some initial efforts to represent this complex characteristics has been presented in [3]. On the other hand, music (and video) offers new ways to measure the context of interaction and the implicit user feedback. Besides the classical implicit feedback evidences, in the case of music access users can replay or skip part of the whole song they like in a particular moment, can adjust the volume, can interrupt other activities or take information about the song or the artist.

Our goal within QONTEXT is to investigate a methodology for measuring the relevance of music items to a particular interaction context by merging different sources of evidence.

5 Data Mining and Visualization

Data Mining is a wide area of research which involves different tasks such as, for example, clustering, categorization, and regression. The main idea behind all these different facets of data mining is the extraction of useful information from data, information which is often implicit or previously unknown. The question is how to model data and how to integrate the outcome of the analyses with visualization components which may help researchers to validate their models. During the last decade, the research area of Quantum Clustering (QC) has given a significant contribution in terms of "non-classical" approaches with efficient clustering algorithms which take advantage of the Quantum Mechanics principles. The work presented by [4] shows how to find the center of the clusters of the dataset calculating the minimum of the potential function $V(\mathbf{x})$ (where \mathbf{x} is a data point) which is based on the definition the *Schrödinger equation*. [5] discuss how to speed up a selection of "classical" clustering algorithms by quantizing some of their parts, and they also suggest that the same paradigm could also be applied to other problems such as dimensionality reduction and training a classifier. Whereas the above cited contributions concentrates on the computational issues of clustering, our research will focus on IAR issues of clustering because there are no results in this area to our knowledge.

Therefore, within QONTEXT, the aim is to explore a more user oriented version of QC. In particular, we want to study the interaction between the contexts of the search process and the clusters, and how the selection of one (or more) cluster by the user can generate a new spaces by means of visual representations of the clusters.

6 Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement N. 247590.

References

- 1. van Rijsbergen, C.J.: The geometry of information retrieval. Cambridge University Press, UK (2004)
- Song, D., Lalmas, M., van Rijsbergen, C.J., Frommholz, I., Piwowarski, B., Wang, J., Zhang, P., Zuccon, G., Arafat, P.B.S., Azzopardi, L., Di Buccio, E., Huertas-Rosero, A., Hou, Y., Melucci, M., Rüger, S.: How quantum theory is developing the field of information retrieval. In: QI'10: Proceedings of the 4th Quantum Interaction Symposium. (2010)
- Bu, J., Chen, C., Wang, C., Wu, H., Zhang, L., He, X.: Music recommendation by unified hypergraph: combining social media information and music content. In: Proceedings of ACM Multimedia. (2010) 391–400
- Horn, D., Gottlieb, A.: Algorithm for data clustering in pattern recognition problems based on quantum mechanics. Phys. Rev. Lett. 88 (2001) 018702
- Aïmeur, E., Brassard, G., Gambs, S.: Quantum clustering algorithms. In: Proceedings of the 24th international conference on Machine learning. ICML '07, New York, NY, USA, ACM (2007) 1–8
- Dahabiah, A., Puentes, J., Solaiman, B.: Possibilistic similarity estimation and visualization. In: Proceedings of the 2nd International Conference on Theory of Information Retrieval: Advances in Information Retrieval Theory. ICTIR '09, Berlin, Heidelberg, Springer-Verlag (2009) 273–280
- 7. Weinstein, M., Horn, D.: Dynamic quantum clustering: A method for visual exploration of structures in data. Phys. Rev. E 80 (2009) 066117