Document Ranking with Quantum Probabilities

Guido Zuccon

Australian E-Health Research Centre

CSIRO

Herston, QLD 4029

Australia

guido.zuccon@csiro.au

2012

Abstract

In this thesis we investigate the use of quantum probability theory for ranking documents. Quantum probability theory is used to estimate the probability of relevance of a document given a user's query. We posit that quantum probability theory can lead to a better estimation of the probability of a document being relevant to a user's query than the common IR approach, i. e. the Probability Ranking Principle (PRP), which is based upon Kolmogorovian probability theory. Following our hypothesis, we formulate an analogy between the document retrieval scenario and a physical scenario, that of the double slit experiment. Through the analogy, we propose a novel ranking approach, the quantum probability ranking principle (qPRP). Key to our proposal is the presence of quantum interference. Mathematically, this is the statistical deviation between empirical observations and expected values predicted by the Kolmogorovian rule of additivity of probabilities of disjoint events in configurations such that of the double slit experiment. While PRP explicitly assumes that the relevancy of a document is independent of that of other documents, we suggest that qPRP implicitly models interdependent document relevance through quantum interference and thus is suited to those document ranking tasks where the independence assumption fails. Throughout the thesis, we also suggest how quantum interference can be estimated for effective document ranking.

To validate our proposal and to gain more insights about approaches for document ranking, we (1) analyse PRP, qPRP and other ranking approaches, exposing the assumptions underlying their ranking criteria and formulating the conditions for the optimality of the two ranking principles, (2) empirically compare three ranking principles (i. e. PRP, interactive PRP, and qPRP) and two state-of-the-art ranking strategies in two retrieval scenarios, those of ad-hoc retrieval and diversity retrieval, (3) analytically contrast the ranking criteria of the examined approaches, exposing similarities and differences, (4) study the ranking behaviours of approaches alternative to PRP in terms of the kinematics they impose on relevant documents, i. e. by considering the extent and direction of the movements of relevant documents across the ranking recorded when comparing PRP against its alternatives.

Our findings show that the effectiveness of the examined ranking approaches strongly depends upon the evaluation context. In the traditional evaluation context of ad-hoc retrieval, PRP is empirically shown to be better than or comparable to alternative ranking approaches. However, when evaluation contexts that account for interdependent document relevance are examined (i. e. when the relevance of a document is assessed also with respect to other retrieved documents, as it is the case in the diversity retrieval scenario), the use of quantum probability theory and thus of qPRP is shown to improve retrieval and ranking effectiveness over the traditional PRP and alternative ranking strategies, such as Maximal Marginal Relevance, Portfolio theory, and Interactive PRP.

This work represents a significant step forward regarding the use of quantum theory in information retrieval. It demonstrates that the application of quantum theory to problems within information retrieval can lead to improvements both in modelling power and retrieval effectiveness, allowing the constructions of models that capture the complexity of information retrieval situations.

Furthermore, the thesis opens up a number of lines of future research. These include investigating estimations and approximations of quantum interference in qPRP, exploiting complex numbers for the representation of documents and queries, and applying the concepts underlying qPRP to tasks other than document ranking.