Physica A 482 (2017) 627–637

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

A quasi-current representation for information needs inspired by Two-State Vector Formalism

Panpan Wang [a], Yuexian Hou [a,*], Jingfei Li [a], Yazhou Zhang [a], Dawei Song [a,b,*], Wenjie Li [c]

[a] Tianjin Key Laboratory of Cognitive Computing and Application, School of Computer Science and Technology, Tianjin University, 300350, No.135 Yaguan Road, Jinnan District, Tianjin, PR China

[b] The Computing and Communications Department, The Open University, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom

[c] Department of Computing and Communications, The Open University, United Kingdom

HIGHLIGHTS

- A session search model is proposed inspired by Two-State Vector Formalism.
- Users' quasi-current information need is captured by the model.
- Extensive experiments have shown its effectiveness.

ARTICLE INFO

Article history:
Received 8 October 2016
Received in revised form 20 April 2017
Available online 28 April 2017

Keywords:
Information Retrieval
Two-State Vector Formalism
Quantum theory
Session search

ABSTRACT

Recently, a number of quantum theory (QT)-based information retrieval (IR) models have been proposed for modeling session search task that users issue queries continuously in order to describe their evolving information needs (IN). However, the standard formalism of QT cannot provide a complete description for users' current IN in a sense that it does not take the 'future' information into consideration. Therefore, to seek a more proper and complete representation for users' IN, we construct a representation of quasi-current IN inspired by an emerging Two-State Vector Formalism (TSVF). With the enlightenment of the completeness of TSVF, a "two-state vector" derived from the 'future' (the current query) and the 'history' (the previous query) is employed to describe users' quasi-current IN in a more complete way. Extensive experiments are conducted on the session tracks of TREC 2013 & 2014, and show that our model outperforms a series of compared IR models.

1. Introduction

Recently, quantum theory (QT), as an important formalism for modeling Information Retrieval (IR) tasks, has attracted increasing attention. van Rijsbergen, in his seminal book [1], proposed for the first time to employ QT as a unified theoretical formalism for modeling IR tasks. The book showed that major IR models (e.g., logical, probabilistic and vector) can be subsumed by a single mathematical formalism in Hilbert vector space (which can be a complex space). In that book, some notions in IR are translated into analogous notions in QT, such as mapping a document into a **state vector**, regarding each document as a **superposition** of words, and replacing the cosine correlation between query and documents with

* Corresponding authors.
E-mail addresses: yxhou@tju.edu.cn (Y. Hou), dwsong@tju.edu.cn (D. Song).

http://dx.doi.org/10.1016/j.physa.2017.04.145
0378-4371/

inner product. Beyond that, QT can help address some problems for IR tasks [2–4]. Following the pioneering work, a series of Quantum Theory-based IR (QIR) models were developed and motivated by quantum probability framework. A representative was Quantum Language Model (QLM) [5], which was presented to model term dependencies in IR and gained good performance for ad-hoc retrieval tasks. However, QLM was solely targeted on single ad-hoc queries and limited its further application on the dynamic search tasks, e.g., session search. To solve that issue, Li et al. [6] developed an adaptive contextual QLM which utilized a density matrix transformation framework to capture the dynamic information (historical queries and clicked documents) within users' search sessions. Then a session-based QLM [7] was also put forward to divide those interaction information into positive and negative feedback to model the evolution of users' information needs (IN). Later, with the inspiration of "quantum interference", the interactive information in a session was used to construct a new superposed state of a document in the IN space [8]. All these models mainly focused on utilizing some concepts or phenomena of QT to describe users' IN. However, they ignore an important fact that the dynamic evolution of users' IN in a session search is supposed be a Markovian-like process from the perspective of QT.

Under the standard formalism of QT, the evolution of a quantum system is a Markovian-like process [9] in the sense that the current quantum state of a system can be completely described by the result of the last measurement. In this paper, a user's IN is analogous to the quantum state, and a series of issued queries are considered as a series of measurement results. Hence, if we use the standard formalism of QT to model the evolution of a user's IN, it should also be a Markovian-like process, in which the current quantum state is determined by the last query (also called the current query in this paper), and is irrelevant with the earlier queries. Note that the above observation could only make sense in a QT framework. We do not suppose that it is universally valid.

However, it is still argued that the standard QT cannot completely characterize a quantum state in a sense that it does not take the information from the 'future' into consideration [10]. In short, the current query might be a proper description for the state of users' current IN, but it is not complete. Note that the 'complete quantum description' corresponds to the completeness under Aharonov, Bergmann, and Lebowitz (ABL) principle mentioned in [9]. To seek a proper representation for users' IN, we construct a quasi-current IN, which contains more complete information than the real current IN, inspired by an emerging Two-State Vector Formalism (TSVF).

Different from the standard formalism of QT, TSVF equips a time-symmetric formulation for QT, which the current system is described by a two-state vector that contains a backward-evolving vector (named post-selected vector) defined by the results of measurement performed on this system in future and a forward-evolving vector (named pre-selected vector). Some evidences have pointed out that such a formalism can provide more complete information than standard QT [11–13]. For example, the result of a measurement of σ_x, σ_y, σ_z performed on a spin-$\frac{1}{2}$ particle at a given time cannot be inferred under the standard QT prescriptions due to the three non-commuting measurements. However, those results can be ascertained with probability 1 under TSVF framework [14]. Currently, TSVF is mainly applied to make interpretations for some unusual quantum phenomena [15,16], such as, Three-boxes paradox and Cheshire Cat. In this article, we apply TSVF to IR tasks for the first time and further investigate how much necessary information we should use to represent the quasi-current IN for QIR models.

According to TSVF, it is incomplete to model users' current IN only by the current query (the query that is not retrieved by search engine yet). To construct a more complete description, our quasi-current IN should be modeled by the historical IN and the future IN. We choose the previous query to represent the historical IN since the earlier queries and interactions in search session are actually deviated from the real IN in accordance with the quantum philosophy. Taking session 2 in 2013 Session Track as an example, the current query "where to buy scooters" tells us that users' current information needs are finding a place to buy scooters. And the previous query "scooter stores" expresses similar meanings. However, the third last query is "scooter price" which is obviously another subtopic about scooter. If the current query and the previous query are combined together to represent users' current IN, the weight of these key words will be increased during retrieving the relevant documents. If the third last query is added, some documents containing the word 'price' will be also retrieved and further disturb user searching for right answers. Therefore, the current query and the previous query combined together can provide a more proper and complete description for users' current IN. Note that the amount of information in our work is not a statistical or information-theoretical concept, but rather an empirical consideration about a description of the past and future information needs of the user. To this end, we formalized a new target function for original QLM [5] inspired by TSVF and obtained a reliable representation for current information needs as density matrices by maximizing the formalized target function. Due to the maximum likelihood estimation methods in QLM cannot ensure global convergence, we replace the original training algorithm with *DilutedRρR* [17]. Extensive experiments are conducted on the session tracks of TREC 2013 and TREC 2014, which show that our model outperforms a series of compared IR models.

2. Related work

van Rijsbergen (2004) provided a mathematical framework based on quantum theory (QT) for the foundations of IR [1]. Using this framework, a document could be represented as a vector in Hilbert space, and the document's relevance could be described by a Hermitian operator. Except that all the usual QT notions, such as superposition state and observation, had their IR-theoretic analogues, the standard QT can be applied to address problems in IR, such as pseudo-relevance feedback and relevance feedback.