



An adaptive contextual quantum language model



Jingfei Li^a, Peng Zhang^{a,*}, Dawei Song^{a,b,**}, Yuexian Hou^a

^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

HIGHLIGHTS

- An adaptive quantum language model for Information retrieval is proposed.
- Users' dynamic information need is captured by the model.
- Extensive experiments have shown its effectiveness.

ARTICLE INFO

Article history:

Received 1 September 2015

Received in revised form 22 January 2016

Available online 28 March 2016

Keywords:

Quantum language model

Density matrix transformation

Session search

Query change information

ABSTRACT

User interactions in search system represent a rich source of implicit knowledge about the user's cognitive state and information need that continuously evolves over time. Despite massive efforts that have been made to exploiting and incorporating this implicit knowledge in information retrieval, it is still a challenge to effectively capture the term dependencies and the user's dynamic information need (reflected by query modifications) in the context of user interaction. To tackle these issues, motivated by the recent Quantum Language Model (QLM), we develop a QLM based retrieval model for session search, which naturally incorporates the complex term dependencies occurring in user's historical queries and clicked documents with density matrices. In order to capture the dynamic information within users' search session, we propose a density matrix transformation framework and further develop an adaptive QLM ranking model. Extensive comparative experiments show the effectiveness of our session quantum language models.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Searching over the Internet and other online data repositories has become a preferred way for people to find relevant information and acquire useful knowledge in our daily life. To fulfill an information need, the users information seeking process is often exploratory, involving continuous interaction with information and dynamic refinement of input queries within a search session. The traditional query–response mode of search (i.e., ad-hoc retrieval) has turned out to be insufficient. Instead, modeling session search is becoming an important yet challenging task. User interactions in search sessions represent a rich source of implicit knowledge about the user's cognitive state and information need that constantly evolves over time. Effectively explicating and incorporating such knowledge in a search system would potentially improve search accuracy and user experience. Despite its potential application values, there still exist some big challenges for this

* Corresponding author.

** Corresponding author at: The Computing and Communications Department, The Open University, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom.

E-mail addresses: pzhang@tju.edu.cn (P. Zhang), dwsong@tju.edu.cn, dawei.song@open.ac.uk (D. Song).

new retrieval mode. For example, modeling term dependencies and capturing the dynamic query change in the context of user interactions. In this paper, we aim at addressing these challenges by presenting two session models inspired by the quantum language model.

The traditional unigram Language Model (LM) for IR is a widely used and robust model for ad-hoc retrieval. It can also be applied in the modeling of the contextual IR which is similar to the session search. However, it does not consider the term dependency in queries which is important to disambiguate the queries. For example, the query LED is ambiguous, which can refer to the Light Emitting Diode (LED) or Latex Editor (LEd). When a user issues a query “led”, the search engine may “misunderstand” the real meaning if the search engine only observe the query term “led”. However, if the search engine observes other words (e.g., semi-conductor, electron, and photon) co-occurring with “led” in a document frequently, we may have high probability to predict the “led” means the Light Emitting Diode (LED).

There have been some IR models which attempted to incorporate the term dependencies, such as the n-gram (e.g., bi-grams and trigrams) language models, extended VSM (Vector Space Model), and MRF (Markov Random Field) models. However, they regard the term dependencies as additional dimensions and fail to model the inter-relationship among components (e.g., single words) of term dependencies. Thus some important information may be lost. For example, traditional dependency models regard “information retrieval” as a special term, the importance of which is weighted as whole with statistics like approaches, while do not quantify the “inner dependency” (inner relationship) between “information” and “retrieval”.

Recently, a Quantum Language Model (QLM) [1] has been proposed for the ad hoc retrieval task, and can model the term dependency in a principled way. In QLM, both queries and documents are represented as a list of projectors that are corresponding to single terms or compound dependencies (i.e., consisting of the internal relations among two or more terms). Projectors are treated as the quantum elementary events, sampled from the quantum probability space, i.e. a Hilbert space. The compound dependency is a superposition event which is a special kind of projector. Over the list of projectors for a query (or a document), there exists a quantum probability distribution encapsulated in a density matrix, known as a Quantum Language Model for the query (or the document). One can utilize an EM-based training algorithm to estimate quantum language models by maximizing a likelihood function. The estimated quantum language models for queries or documents can not only model the importance of single terms with the diagonal elements of the density matrix, but also quantify the inner relationship between components of each phrase or term dependencies with the non-diagonal elements. Let us now look at the phrase “information retrieval” as an example. In traditional LM, the phrase is regarded as a variable often with the frequency as its value, while in the QLM, the phrase is represented as a 2-order density matrix. The values of diagonal elements denote the weights of each single word, and the non-diagonal elements denote the relationship between two single words. It can be observed that the matrix representation of the phrase “information retrieval” contains more information about the inner dependency of the phrase “information retrieval”. Furthermore, the ranking function for a document is the VN-divergence [1,2] between query and document QLMs.

Motivated by the sound theory of QLM, in this paper we propose a session search approach, denoted as Contextual QLM (C-QLM), which naturally incorporates the complex term dependencies occurring in user’s historical queries and clicked documents. C-QLM utilizes user’s historical queries, clicked documents and the pseudo-relevance documents to obtain a reliable representation of user’s hidden information need as a contextual query QLM. It quantifies the internal relationship among components of the term dependencies by encapsulating the QLM into a density matrix which is trained by maximizing a likelihood function of a document representation as a sequence of projectors. Projectors can capture more compound dependency among terms. The C-QLM significantly improves the original ranked results returned by the search engine.

However, the C-QLM does not model the dynamics of users’ information need. Thus it fails to capture the evolution information when users are changing their queries to find final search results. To deal with this problem, we further improve the C-QLM model and propose an Adaptive Contextual QLM (AC-QLM) to capture the query reformulation information when the user is completing a specific search task. To do this, we assume that there exists an ideal density matrix which can well represent user’s real-time information need perfectly. However, it is impossible to obtain this “ideal” density matrix, we can only approach the ideal density matrix as closely as possible. To this end, we propose to transform the density matrix for contextual query QLM by incorporating query reformulation information within a search session. An adaptive algorithm is developed to approximate the density matrix transformation framework. From the experimental results, we find that the AC-QLM improved the ranking of more relevant documents, compared with the C-QLM. Although there have been some attempts to use the query change information in session search task [3,4], they do not consider term dependencies in query representation.

The rest of this paper is organized as follows. Section 2 presents a brief review of the related work. Section 3 gives a detailed introduction to the quantum language model. The QLM-based session search models are proposed in Section 4. In Section 5, we report the empirical experiments. Section 6 concludes the paper and points out some future research directions.

2. Related work

There are two lines of related work, i.e. search personalization and quantum theory inspired information retrieval.

Download English Version:

<https://daneshyari.com/en/article/7377817>

Download Persian Version:

<https://daneshyari.com/article/7377817>

[Daneshyari.com](https://daneshyari.com)