

# Modeling Quantum Entanglements in Quantum Language Models\*

Mengjiao Xie<sup>1</sup>, Yuexian Hou<sup>1</sup>, Peng Zhang<sup>1</sup>, Jingfei Li<sup>1</sup>, Wenjie Li<sup>2</sup>, Dawei Song<sup>1,3</sup>

<sup>1</sup>School of Computer Science and Technology, Tianjin University, China

<sup>2</sup>Department of Computing, The Hong Kong Polytechnic University, Hong Kong

<sup>3</sup>Department of Computing and Communications, The Open University, UK

{mjxie, yxhou, pzhang, jingfeili, dwsong}@tju.edu.cn, cswjli@comp.polyu.edu.hk

## Abstract

Recently, a Quantum Language Model (QLM) was proposed to model term dependencies upon Quantum Theory (QT) framework and successively applied in Information Retrieval (IR). Nevertheless, QLM's dependency is based on co-occurrences of terms and has not yet taken into account the Quantum Entanglement (QE), which is a key quantum concept and has a significant cognitive implication. In QT, an entangled state can provide a more complete description for the nature of realities, and determine intrinsic correlations of considered objects globally, rather than those co-occurrences on the surface. It is, however, a real challenge to decide and measure QE using the classical statistics of texts in a post-measurement configuration. In order to circumvent this problem, we theoretically prove the connection between QE and statistically Unconditional Pure Dependence (UPD). Since UPD has an implementable deciding algorithm, we can in turn characterize QE by extracting the UPD patterns from texts. This leads to a measurable QE, based on which we further advance the existing QLM framework. We empirically compare our model with related models, and the results demonstrate the effectiveness of our model.

## 1 Introduction

Quantum theory (QT) has a profound impact on multidisciplinary studies, such as Physics, Philosophy, Psychology and Information Retrieval (IR) [Heisenberg and Northrop, 1958; Busemeyer and Bruza, 2012; Melucci and van Rijsbergen, 2011; Song *et al.*, 2010; Sordoni *et al.*, 2014]. In a seminal book [van Rijsbergen, 2004], quantum theory was for the first time regarded as “a kind of formal language that can be used for describing objects and processes in IR”. Following this pioneering work, the Quantum Probability Ranking Principle (QPRP) was proposed to rank interdependent documents based on the quantum interference [Zuccon and Az-zopardi, 2010]. Inspired by the photon polarization, a Quantum Measurement inspired Ranking model (QMR) was in-

roduced [Zhao *et al.*, 2011]. Although the QPRP and the QMR achieved better performances, the probabilities in both methods are not actually calculated according to the QT formulations.

Piwowski *et al.* [2010] explored to express queries as density matrices and documents as subspaces. However, this method did not bring about the excellent performance as expected. Later, a Quantum Language Model (QLM) was proposed to model term dependencies based on the quantum theory and outperformed the conventional IR model [Sordoni *et al.*, 2013]. However, QLM models the term dependency using traditional statistics corresponding to co-occurrences of terms, thus lacks of an explicit modeling of the Quantum Entanglement (QE), which is a key quantum concept and has a significant cognitive implication.

As a quantum model, e.g., QLM, it is inconsistent if it cannot express the quantum entanglement in itself. An entangled state can give a more complete description for the nature of global realities, in which the intrinsic correlations among objects are globally determined [Atmanspacher, 2012]. The quantum entanglement is relevant to cognition since it is natural to model the holistic and instantaneous interaction in a cognitive subject [Atmanspacher, 2012]. Technically speaking, a system is entangled, if and only if its state cannot be expressed as a tensor product of its sub-systems. However, in IR practice, it is challenging to measure QE using the classical statistics of texts in a post-measurement configuration, where the state of an object (e.g., the mind of a man) has collapsed to certain observables (e.g., texts).

We resort to the recently proposed Unconditional Pure Dependence (UPD) [Hou *et al.*, 2013] to deal with the above-mentioned problem. It is proved that in a post-measurement configuration, the statistical dependence is equivalent to entanglements [Hou and Song, 2009]. In this paper, we reformulate the above statistical dependence as the so-called UPD. Hence we can indeed prove the equivalence relationship between UPD and QE in a post-measurement configuration. Since Hou *et al.* [2013] have proposed an efficient algorithm to decide, in a sufficient sense, whether a set of variables is UPD, it turns out that we can characterize QE by extracting the UPD patterns from textual datasets.

Moreover, Hou *et al.* [2013] emphasized that a UPD pattern is not formed by accidental co-occurrences of the words in the sense that its joint distribution cannot be uncondition-

\*Corresponding authors: Yuexian Hou and Peng Zhang.

ally factorized. This is coincident with what Ludwig Wittgenstein said: “For a large class of cases of the employment of the word ‘meaning’ . . . . . the meaning of a word is its use in the language” [Wittgenstein, 1953]. Wittgenstein defined the word’s meaning as the use of the word. Here, the “use” can be characterized by the relationship of the non-accidental co-occurrences of words to a large extent. Since the non-accidental co-occurrence relationship can be characterized by UPD, which is different from the dependence modeled by QLM, those UPD patterns can further help QLM to model meaningful patterns of words, unlike the unprincipled setting in QLM where accidental co-occurrences of terms are also considered. Through integrating UPD into the QLM’s training algorithm, QLM thus naturally takes into account QE, leading to a more consistent quantum IR model.

The rest of this paper is laid out as follows. We first introduce the preliminaries of QT in Section 2. In Section 3, we discuss QE. Then we illustrate the relationship between QE and UPD in Section 4, where we also provide the equivalence proof between QE and UPD for the sake of understandability. We propose to advance QLM with UPD in Section 5. In Section 6, we present the experimental results of our model. Section 7 concludes this work and points out future research directions.

## 2 Preliminaries of Quantum Theory

In quantum theory [Nielsen and Chuang, 2010], a probabilistic space is defined as a Hilbert space  $\mathbb{H}^n$ , which is an abstract vector space possessing the structure of the inner product. For convenience, we restrict ourselves to a real space  $\mathbb{R}^n$ . With the Dirac’s notation, a unit column vector  $\varphi \in \mathbb{R}^n$  corresponding to a quantum state can be expressed as the ket  $|\varphi\rangle$  and its transpose  $\varphi^T$  is the bra  $\langle\varphi|$ . The inner product of  $\varphi$  and  $\phi$  is expressed as  $\langle\varphi|\phi\rangle$ . Onto the direction  $|\varphi\rangle$ , the corresponding measurement projector is  $|\varphi\rangle\langle\varphi| = \varphi\varphi^T$ .  $|\varphi\rangle\langle\varphi|$  can also represent a quantum state, and hence becomes the density matrix. According to the measurement postulate, the expectation value of the measurement on a density matrix  $\rho$  can be calculated by  $\mathbf{E}(H) = \sum_m mp(m) = \text{tr}(\rho H)$ ,

where  $H$  is a measurement Hermitian with a spectral decomposition. That is,  $H$  is equal to its own conjugate transpose and  $H = \sum_m mH_m$ , where  $H_m$  is the projector onto the eigenspace of  $H$  with eigenvalue  $m$ . In this paper,  $H$  is simply defined as a single projector (See Formula (6)). It turns out that  $\text{tr}(\rho H)$  gives the probability that the measurement result corresponding to this projector occurs. All these probability terms can form a quantum probability measure on all possible projectors since Gleason’s Theorem [Gleason, 1957] proves a bijective relationship between the density matrix  $\rho$  and the quantum probability measure  $\mu$ , which is determined by  $\mu_\rho(|\varphi\rangle\langle\varphi|) = \text{tr}(\rho|\varphi\rangle\langle\varphi|)$ .

## 3 Quantum Entanglement and Its Cognitive Implications

In quantum theory, a qubit can be completely described by a state vector. Importantly, a qubit can be in a superposition

state, i.e.,  $a|0\rangle + b|1\rangle$ , where  $|a|^2 + |b|^2 = 1$ . The special states  $|0\rangle$  and  $|1\rangle$  forming an orthonormal basis are computational basis states. When a qubit is measured, the superposition state will collapse into one of the base states with certain probability, i.e., the squared norm of the corresponding coefficient of the base. Specially, the entangled state is a special case of the superposition state involving multiple qubits.

The Quantum Entanglement (QE) can be formally defined as follows:

**Definition 1.** (QE): Let  $A$  be an  $n$ -qubit system in a state  $|\phi_A\rangle$  and  $\{A_1, A_2\}$  be a partition of  $A$ , where two disjoint parts  $A_1$  and  $A_2$  have  $0 < k < n$  qubits and  $n - k$  qubits, respectively.  $A$  is entangled iff. there does NOT exist any tensor product decomposition of  $|\phi_A\rangle$  such that  $|\phi_A\rangle = |\phi_{A_1}\rangle \otimes |\phi_{A_2}\rangle$ , where  $|\phi_{A_1}\rangle$  and  $|\phi_{A_2}\rangle$  are the states of  $A_1$  and  $A_2$ , respectively.

As an illumination of quantum entanglement, we give an example here. Suppose a state with two qubits is written as  $\sqrt{1/2}(|00\rangle + |11\rangle)$ , which is the well-known Bell entangled state, also called the EPR pair [Rieffel and Polak, 2000]. From the perspective of the quantum measurement, it is obvious that any qubit in EPR pair will be completely determined by the measurement result of the other qubit in this pair.

In the following, we describe the cognitive implications of quantum entanglement from its philosophical background and a specific example, i.e., the surprise examination paradox [Chow, 1998; Kritchman and Raz, 2010]. From a philosophical point of view, Pauli and Jung made an analogy that “the change from the unconscious state to a conscious state can be seen as the converting from an entangled state of a system to a measured state” [Atmanspacher, 2012]. They argued that the recognition of the unconscious state is akin to the quantum measurement on the entangled state. Atmanspacher [2012] stated that “measurements can be viewed as an intervention decomposing a system constituting an inseparable whole into locally separate parts”. Therefore, in the light of measurement results, we may speculate the knowledge about the holistic state before the measurement.

As we usually directly get the semantic knowledge from texts or propositions in our life, we develop the idea of Pauli and Jung as illustrated in Figure 1. That is, the consciousness, such as the mind and the meaning, is mostly manifested in the form of texts or propositions by thinking logically. After the quantum-like measurement, the consciousness collapses to texts or propositions, which we observe in reality. As illuminated by the above analogies, we can conclude that the conversion of a meaning system from an entangled state to a measured state corresponds with the transformation of a meaning system from a global and implicit state to a local and explicit state due to measurements. Here the measurement results are often texts or propositions.

To further clarify the meaning of the quantum entanglement, we turn to discuss a notable paradox, namely the surprise examination paradox. Here, we give a simplified version: a teacher says to students that there will be an examination in the class on Monday or Wednesday next week and you will not infer the specific date for the examination. If the

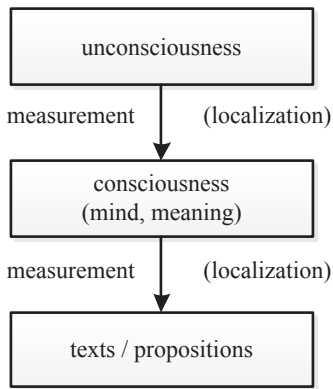


Figure 1: Transformations among the unconsciousness, the consciousness, texts and propositions. The localization represents that the globally implicit meaning is explicitly expressed by texts or propositions.

teacher’s claims are formulated under the framework of the Classical Propositional Logic (CPL) system, it can deduce a contradiction. But interestingly, the paradox happens since the examination could be actually held in a class and no student knows the specific date before it happens, which indeed meets the teacher’s claims.

This reminds us of Wittgenstein’s statement: there are truths of silence, when spoken, no longer true anymore [Wittgenstein, 1961]. Following Wittgenstein’s point, the phenomenon of the examination paradox can be understood intuitively. That is, once compatible facts are formalized as propositions and regarded as the premises of the reasoning in the CPL system, conflicts may happen. Actually, Gödel’s second incompleteness theorem [Gödel, 1931] theoretically explains this phenomenon. That is, the system explicitly introduces a meta proposition on the unprovability of the system, which claims that students cannot infer the specific date for the examination, so that the consistency of the system is broken. We note that the above observations can also be understood under the framework of quantum theory. Specifically, in quantum cognition, there exist holistic truths in the form of entangled states, which cannot be explicitly expressed by propositions or texts. The way to express these holistic truths in a (formal or natural) language can be viewed as the quantum measurement. Although the results of quantum measurements locally reflect holistic truths, a localization (i.e., explicit expressions via propositions or texts) may break the consistency of the expression system.

The examination paradox reveals the limitations of the CPL system. It is thus necessary to seek for a formal and replaceable system to describe the states of entangled objects. In quantum theory, the examination paradox can be simply and completely described with two entangled qubits. Formally, for the first qubit, 0 and 1 represent the examination is held on Monday or Wednesday, respectively. For the second qubit, 0 and 1 represent the students cannot or can infer the specific date for the examination, respectively. Now, the student’s cognitive state before the examination can be modeled

as  $a|00\rangle + b|11\rangle$ , an entangled state, where  $|a|^2 + |b|^2 = 1$ . The class on Monday can be seen as a measurement on the first qubit, which makes the resulting state become localized propositions. Specifically, if the examination is on Monday, the entangled state collapses to  $|00\rangle$ , which means that the examination is held on Monday and students cannot infer the date for the examination. Otherwise, the student’s cognitive state collapses to  $|11\rangle$ . No matter what the outcome is, the student’s cognitive state changes substantially owing to the quantum measurement. This is in sharp contrast with the CPL system, in which the procedures of the reasoning do not produce any new knowledge in nature, since all logical conclusions are actually implied by the logical premises. Therefore, the state of the CPL system is essentially static. On the other hand, the quantum state can characterize not only whether the proposition comes into being or not, but also the probability of the occurrence of the proposition in principle. In this sense, the quantum state may give a more complete description than the CPL system. Main differences between the formalizations of the CPL and the quantum state representation are listed in Table 1.

The above analysis indicates that the quantum state may be more useful and powerful than the CPL system in some contexts. In particular, the quantum entanglement reflects the intrinsic correlations among the potential propositions we concern about. It is these correlations that globally determine the whole student’s cognitive state and the state collapsing induced by measurements. Furthermore, a quantum state may also evolve in accordance with the unitary matrix. Therefore, we can define the QT model as a model that describes the states of objects by, in general, entangled state vectors, and characterizes the localization and evolution of the states by quantum measurements and unitary evolutions, respectively. It is worth emphasizing that, in this framework, the quantum entanglement is a key component since it rules the holistic behaviors of the quantum state.

In IR, our task is to retrieve documents that are relevant to a query topic. The underlying meaning of a document globally determine which topic it belongs to. To a large extent, we can capture the underlying meaning through the high-level meaningful entities/blocks. It is natural that this kind of entities is constituted by a set of intrinsically correlated components in an entangled (meaning) state. In a post-measurement configuration, in which the meaning is fixed by the measurement results, i.e., texts, an entangled state manifests itself as the special statistical dependence among words (see Section 4 for details). For example, a text pattern “Seattle World fair”, in which the words are significantly dependent, implies the whole concept of “world’s fair in Seattle”, i.e., a high-level meaningful entity. As a conclusion, it is important to characterize the quantum entanglement in an IR model.

#### 4 The Relationship between QE and UPD

We expect to model QE in QLM. However, it is usually impossible for us to directly measure a quantum-like system, e.g., the mind of a man, where the quantum measurement occurs in the generation process of texts. In common applications, we have only these texts or statistical data, which corre-

	Formulation of CPL system	Formulation of Quantum system
Representation of system state	Localized propositions	Quantum superposition states
Reasoning of system	Reasoning rules of CPL	Quantum measurements
Temporal property of system	Static	Dynamic

Table 1: Main differences between a CPL system and a Quantum system.

spond to the measurement results of the underlying quantum-like system. While, the statistics of texts comply with classical logic and probability, which are incompatible with pure quantum approaches. Hence we need a criterion to decide the existence of quantum entanglement from the measurement results of a quantum system. To this end, Hou and Song [2009] described that the entanglement corresponds to the statistical dependence between measurement results. In the following, we reformulate the above statistical dependence as the unconditional pure dependence. Hence we can indeed prove the equivalence relationship between the unconditional pure dependence and the quantum entanglement in a post-measurement configuration.

In a post-measurement configuration, the counterpart of QE is Unconditional Pure Dependence (UPD) defined as below. Formally, given a set of binary random variables,  $A = \{W_1, W_2, \dots, W_n\}$ , where  $W_i$  denotes the occurrence ( $W_i = 1$ ) or absence ( $W_i = 0$ ) of the  $i^{\text{th}}$  word. Let  $w_i \in \{0, 1\}$  stand for the value of  $W_i$  and  $p(\mathbf{a})$ ,  $\mathbf{a} = [w_1, w_2, \dots, w_n]^T$  be the joint distribution over  $A$ . Then we have:

**Definition 2.** (UPD): A pattern  $A = \{W_1, W_2, \dots, W_n\}$  forms the UPD pattern iff. the joint probability distribution over  $A$  cannot be unconditionally factorized, i.e., there does NOT exist any  $m$ -partition  $\{A_1, A_2, \dots, A_m; m > 1\}$  of  $A$ , so that  $p(\mathbf{a}) = p(\mathbf{a}_1) \cdot p(\mathbf{a}_2) \cdot \dots \cdot p(\mathbf{a}_m)$ , where  $p(\mathbf{a}_i)$ ,  $i = 1, 2, \dots, m$ , is the joint distribution over  $A_i$ .

Based on Definition 1 and Definition 2, a set of objects in an entangled system or a UPD pattern are both recognized as a whole, which cannot be decomposed by the individual counterparts in the sense that the states of any part are intrinsically correlated with the states of the rest. This intuitive similarity between QE and UPD implies a possible logical connection between them. In the following, we will formally clarify this connection.

To uncover the relation between QE and UPD, we first define the characterizing distribution of a quantum state. Given a quantum state  $|\phi\rangle$ , the characterizing distribution of  $|\phi\rangle$  is the (classical) joint probability distribution indicating the occurrence probability of all possible measurement results on all qubits of  $|\phi\rangle$ . For example, if  $|\phi\rangle = \sqrt{1/2}(|00\rangle + |11\rangle)$ , then the characterizing distribution of  $|\phi\rangle$  is just a discrete joint distribution with  $p_{00} = 1/2$ ,  $p_{01} = 0$ ,  $p_{10} = 0$  and  $p_{11} = 1/2$ .

**Proposition 1.** Let  $A$  be an  $n$ -qubit system in a state  $|\phi_A\rangle$ .  $A$  is entangled iff. the characterizing distribution of  $|\phi_A\rangle$  is unconditional pure dependent.

*Proof.* To prove the sufficiency, we assume that  $A$  is not entangled, i.e.,  $|\phi_A\rangle$  has a tensor product composition  $|\phi_A\rangle = |\phi_{A_1}\rangle \otimes |\phi_{A_2}\rangle$ , where  $|\phi_{A_1}\rangle$  and  $|\phi_{A_2}\rangle$  corresponds to the

states of two disjoint parts of  $A$ , i.e.,  $A_1$  and  $A_2$ , respectively. Note that  $\{A_1, A_2\}$  should be a partition of  $A$ . Without loss of generality, we suppose that  $|\phi_{A_1}\rangle$  includes the first  $k$  qubits of  $|\phi_A\rangle$  and  $|\phi_{A_2}\rangle$  includes the last  $(n - k)$  qubits of  $|\phi_A\rangle$ . Denote  $|\phi_A\rangle = z_{0\dots 0}|0\dots 0\rangle + \dots + z_{1\dots 1}|1\dots 1\rangle$ ,  $|\phi_{A_1}\rangle = x_{0\dots 0}|0\dots 0\rangle + \dots + x_{1\dots 1}|1\dots 1\rangle$  and  $|\phi_{A_2}\rangle = y_{0\dots 0}|0\dots 0\rangle + \dots + y_{1\dots 1}|1\dots 1\rangle$ . Note that the dimensionalities of  $|\phi_A\rangle$ ,  $|\phi_{A_1}\rangle$  and  $|\phi_{A_2}\rangle$  are different. Substitute these representations into the tensor product composition of  $|\phi_A\rangle$ , it is easy to check that the following relations holds:

$$z_{i_1 \dots i_n} = x_{i_1 \dots i_k} y_{i_{(k+1)} \dots i_n}, \quad (1)$$

for any  $i_1 \dots i_n$ , where  $\forall i_j \in \{0, 1\}$ . The Formula (1) exactly implies that the characterizing distribution of  $|\phi_A\rangle$  is not unconditional pure dependent. The sufficiency follows.

To prove the necessity, we assume that the characterizing distribution of  $|\phi_A\rangle$  is not unconditional pure dependent. That is, the characterizing distribution of  $|\phi_A\rangle$  has a factorization. It is easy to check that a factorization of the characterizing distribution of  $|\phi_A\rangle$  entails a tensor product composition of  $|\phi_A\rangle$  as depicted in Formula (1). It turns out that the  $|\phi_A\rangle$  is not entangled. The necessity follows.  $\square$

Based on Proposition 1, QE is equivalent to UPD in a post-measurement configuration. Hence we can decide the existence of QE via UPD, and furthermore use the measure of UPD as a quantitative index of QE.

## 5 A Quantum IR model with Quantum Entanglements

Quantum Language Model (QLM) is a recently proposed quantum-like method to retrieve documents given query topics. To make QLM with more quantum properties, we propose to enhance QLM by characterizing the quantum entanglement in QLM. Based on Proposition 1, the issue of characterizing the quantum entanglement converts to identify the UPD patterns from texts. Since deciding the UPD is NP-hard in general, Hou [2013] developed a UPD pattern extraction method based on Information Geometry (IG), which can efficiently decide the UPD in a sufficient sense. With this method, an enriched Quantum IR model with Quantum Entanglements (QQE) can be naturally realized.

Similar to QLM, we represent documents by a series of projectors. However, the way to build projectors is different from the original QLM since we use a different procedure to choose dependence patterns, which thereby results in different actual representations of documents and queries. In our QQE model, the procedures of building the sequence of projectors for the document are as follows.

Firstly, for a specific query, we extract UPD patterns using a Log Likelihood Ratio Test (LLRT), which is designed under

a principled IG framework [Hou *et al.*, 2013]. The testing statistic  $\rho$  [Nakahara and Amari, 2002] of LLRT is derived as follows:

$$\rho \approx Ng_{dd}\hat{\theta}_{12\dots n}^2, \quad (2)$$

where  $N$  is the sample numbers, the values of  $g_{dd}$  and  $\theta$ -coordinate can be calculated by the method given in [Hou *et al.*, 2013]. The degree that a pattern is unconditional pure dependent in terms of the statistic  $\rho$  of LLRT is estimated by the confidence level  $\pi$ .

$$\pi = \text{cdf}_{\chi^2(1)}(\rho), \quad (3)$$

where  $\text{cdf}$  denotes the cumulative distribution function of  $\chi^2(1)$ . For instance, given a query  $Q$  with  $n$  words,  $Q = \{w_j; j = 1, \dots, n\}$  and a pattern  $\{w_1, w_2\}$  with computed  $\rho=5.024$ , we can guarantee that this pattern is a UPD pattern with probability 0.95. For more details of the extracting method, please refer to Section 5 in [Hou *et al.*, 2013].

Secondly, we represent uni-terms and UPD patterns as projectors based on the method proposed in [Sordoni *et al.*, 2013]. For the query  $Q$ , suppose the set  $\{|e_{w_j}\rangle\rangle$  forms an orthonormal basis. The projector for a uni-term  $w_j$  is below:

$$m(w_j) = |e_{w_j}\rangle\rangle\langle e_{w_j}|, \quad w_j \in Q. \quad (4)$$

The projector for a UPD pattern  $K = \{w_i; i = 1, \dots, h\}$  is as follows:

$$m(K) = |e_k\rangle\rangle\langle e_k|, \quad |e_k\rangle\rangle = \sum_{i=1}^h \sigma_i |e_{w_i}\rangle\rangle, \quad \sum_{i=1}^h \sigma_i^2 = 1, \quad (5)$$

where  $\sigma_i$  can be assigned to the uniform ( $\sigma_i = \sqrt{1/n}$ ) weight or the normalized inverse document frequency (idf) weight.

Thirdly, for each occurrence of uni-terms and UPD patterns in unordered fixed windows (with length  $L$ ) of a document, we add their projectors to the sequence of projectors for the document. Specifically, we design Algorithm 1 to build the sequence of projectors for a document  $D$ . Note that  $\#\text{uw}L(w_j, D)$  or  $\#\text{uw}L(K, D)$  is an Indri expression which returns the occurrence times of  $w_j$  or  $K$  in unordered windows (with length  $L$ ) of  $D$ .

---

**Algorithm 1** Build the sequence of projectors  $S_D$  for  $D$ .

---

**Require:**  $Q = \{w_j; j = 1, \dots, n\}$ ;  $D$ ;  $U$ : the UPD patterns for  $Q$ .

```

1: Initialize  $S_D \leftarrow \emptyset$ ;
2: for each  $w_j$  in  $Q$  do
3:   for  $i = 1$  to  $\#\text{uw}L(w_j, D)$  do
4:     add  $m(w_j)$  to  $S_D$ ;
5:   end for
6: end for
7: for each pattern  $K$  in the power set  $\mathcal{P}(Q) - \{\emptyset\}$  do
8:   if  $K \subset U$  then
9:     for  $i = 1$  to  $\#\text{uw}L(K, D)$  do
10:      add  $m(K)$  to  $S_D$ ;
11:     end for
12:   end if
13: end for
14: return  $S_D$ 

```

---

After obtaining the sequence of projectors for documents, we follow the same training method as QLM [Sordoni *et al.*, 2013] to estimate density matrices. Specifically, given a sequence of projectors  $S_D = m_1, \dots, m_N$  for a document  $D$ , in QLM, density matrices are estimated by:

$$\max_{\rho} \log \prod_{i=1}^N \text{tr}(\rho m_i). \quad (6)$$

The  $R\rho R$  algorithm [Lvovsky, 2004] is used for solving the above maximization problem. The Dirichlet smoothing method is used to smooth the density matrices. In Formula (6), the projectors consist of the mappings of uni-terms and UPD patterns. This is different from QLM, in which the uni-terms and all possible combinations of uni-terms are modeled as projectors. In this sense, our model can reduce the time complexity when we train the density matrix.

Finally, the negative quantum relative entropy of the query density matrix  $\rho_Q$  to the document density matrix  $\rho_D$  is considered as the scoring function:

$$-S(\rho_Q \parallel \rho_D) = -\text{tr}(\rho_Q(\log \rho_Q - \log \rho_D)) \\ \stackrel{\text{rank}}{=} \text{tr}(\rho_Q \log \rho_D). \quad (7)$$

## 6 Experimental Results and Analyses

To verify the effectiveness of our model, we conduct experiments on five collections in the documents re-ranking task. Table 2 describes the collections in detail. Only the title queries are considered in the experiments. The collections have been stemmed with the Porter stemmer and processed with the standard stop word list when using the Lemur 4.12 for building index. For all the models, we apply the Dirichlet smoothing method [Zhai, 2008] in which the parameter  $\mu$  is set to be a typical value 2500. A pool of  $N(N=80,100)$  documents retrieved by the unigram model is used for re-ranking, while we obtain similar observations for other  $N$ 's settings. Mean Average Precision (MAP) is adopted as the metric to evaluate the retrieval performance. We ignore the queries whose relevant documents are not given in the standard results. The Student's test is used as the statistical significance test method, where the level is 0.05. The unconditional pure dependence patterns with at most three orders are taken into account, in order to reduce the computational cost. The length of the unordered fixed window in Algorithm 1 is set to be:  $L = 4 \cdot |K|$ , where  $|K|$  is the numbers of terms appearing in the UPD pattern  $K$  [Metzler and Croft, 2005].

Name	#docs	Topic
FR	25960	151-200
ZIFF	75180	151-200
SJM	90257	101-150
DOE	226087	151-200
WT10G	1692096	501-550

Table 2: A summary of collections and topics.

MAP% (+chg%)	topN=80					topN=100				
	FR	ZIFF	SJM	DOE	WT10G	FR	ZIFF	SJM	DOE	WT10G
QLM_uni	24.85	24.88	17.39	13.69	13.70	25.07	24.97	17.80	13.97	14.73
QQE_uni	27.06 <sup>α</sup> (8.89)	26.41 <sup>α</sup> (6.15)	18.27 <sup>α</sup> (5.06)	14.25 (4.09)	14.04 (2.48)	27.31 <sup>α</sup> (8.93)	26.47 <sup>α</sup> (6.01)	18.75 <sup>α</sup> (5.34)	14.53 (4.01)	15.05 (2.17)
QLM_idf	25.20	24.89	17.34	13.92	13.77	25.46	25.00	17.73	14.19	14.71
QQE_idf	28.04 <sup>β</sup> (11.27)	26.03 (4.58)	18.24 <sup>β</sup> (5.19)	14.53 (4.38)	14.05 (2.03)	28.31 <sup>β</sup> (11.19)	26.13 (4.52)	18.66 <sup>β</sup> (5.25)	14.81 (4.37)	15.07 (2.45)

Table 3: Results for QQE and QLM assigned unifrom (uni) and idf weights respectively. The statistical significant improvements over QLM\_uni and QLM\_idf are marked with  $\alpha$  and  $\beta$  respectively. The values in parentheses in the fourth and the last lines indicate the relative improvements (+%) over QLM\_uni and QLM\_idf, respectively.

MAP% (+chg%)	topN=80					topN=100				
	FR	ZIFF	SJM	DOE	WT10G	FR	ZIFF	SJM	DOE	WT10G
unigram	24.26	21.44	16.96	12.92	12.19	24.57	21.63	17.55	13.25	13.01
MRF_T_U	24.56	21.91	17.16	12.81	13.43	24.85	22.22	17.79	13.14	14.34
MRF_T_UPD	26.11	21.84	17.28	13.78	13.44	26.43	22.13	17.91	14.14	14.39
QQE_idf	28.04 <sup>γ</sup> (7.39)	26.03 <sup>γ</sup> (19.18)	18.24 <sup>γ</sup> (5.56)	14.53 (5.44)	14.05 (4.54)	28.31 <sup>γ</sup> (7.11)	26.13 <sup>γ</sup> (18.08)	18.66 <sup>γ</sup> (4.19)	14.81 (4.74)	15.07 (4.73)

Table 4: Results for unigram, MRF\_T\_U, QQE and MRF\_T\_UPD. The statistical significant improvements over MRF\_T\_UPD are marked by  $\gamma$ . The values in parentheses in the last line indicate relative improvements (+%) over MRF\_T\_UPD.

## 6.1 Our Model (QQE) vs. QLM

As discussed previously, a main advantage of our model QQE described in Section 5 over the original QLM is that QQE characterizes the quantum entanglement by retaining UPD patterns while removing non-UPD ones in choosing projectors of QLM. We report the experimental results in Table 3.

From Table 3, we can see that our model achieves stable improvements over QLM on all the five collections. From the results of the statistical significant test, QQE is significantly effective on FR, ZIFF and SJM collections, in comparison with QLM. We then investigate the lengths of queries in these collections. An observation is that QQE is more effective for the relatively long queries. This is possibly because that there are more possible combinations of uni-terms as dependence patterns for the longer query in the original QLM, but some dependence patterns (especially the non-UPD ones) could introduce noises and may be irrelevant to certain queries. Different weight parameters  $\sigma_i$  (the uniform weight or the idf weight) in Formula (5) also influence the retrieval performance. The results show that QQE\_idf outperforms QQE\_uni slightly on FR, DOE and WT10G collections. It indicates that idf weights seem to be more reasonable than uniform weights.

## 6.2 Our Model (QQE) vs. MRF\_T\_UPD

The Markov Random Field (MRF) retrieval model [Metzler and Croft, 2005] makes use of three features, i.e., uni-terms, sequential terms and unordered terms. The UPD [Hou *et al.*, 2013] was once adopted in the MRF model, for the purpose of refining the sets of sequential terms and unordered terms. Therefore, it is necessary to compare the performance of UPD patterns used in the traditional retrieval model (e.g., MRF)

and the quantum retrieval model (e.g., QLM), respectively. In order for a fair comparison, we only use UPD to refine the sets of unordered terms. In our experiments, the unigram model [Song and Croft, 1999] and the MRF\_T\_U model are involved. The capital T, U and UPD represent that the model uses uni-terms, unordered terms (#uw) and UPD patterns, respectively. The detailed results are presented in Table 4.

Table 4 shows that QQE outperforms the unigram model substantially in the five collections. In comparison with MRF\_T\_UPD, QQE still achieves better performances. Particularly, MRF\_T\_UPD is slightly worse than MRF\_T\_U on ZIFF collection, but QQE\_idf achieves larger improvements over MRF\_T\_U on this collection. These results indicates that it seems to be more useful to apply UPD patterns in a quantum IR model (e.g., QLM). This is possibly because the UPD patterns, which have proved connections with the quantum entanglements, can be naturally modeled in building projectors in QLM.

In summary, two groups of experiments show similar observations on the effectiveness of the proposed QQE. Specifically, the results show that QQE achieves better performances than QLM (a quantum IR model) and MRF\_T\_UPD model (a traditional IR model with UPD patterns).

## 7 Conclusions and Future Works

In this paper, we introduce the quantum entanglement in the field of information retrieval. In order to measure and decide QE, we theoretically prove the equivalence between QE and UPD in a post-measurement configuration. This help us infer the occurrence of QE by extracting UPD patterns. We then enhance a Quantum IR model (i.e., QLM) with Quantum En-

tanglements (QQE) by reserving the UPD patterns in building projectors in QLM. The experimental results show that our model QQE is more effective than the quantum language model and a traditional IR model namely MRF.

Our future work will be focused on analysing the proposed methodology and specific models on some other IR/NLP tasks (e.g., the session search) and investigating the resulted new problems and observations. We will also analyse how to effectively set the value of  $\sigma_i$  (in Formula 5) in a more principled manner. Furthermore, we will develop some other effective representations for documents and queries, using the concept of QE and UPD.

## Acknowledgments

This work is funded in part by the Chinese 863 Program (grant No. 2015AA015403), the Chinese 973 Program (grant No. 2013CB329304 and 2014CB744604), the Chinese NSF Project (grant No. 61272291, 61402324 and 61272265), the Key NSF Project of Chinese Tianjin (grant No. 15JCZDJC31100) and the Major Project of Chinese National Social Science Fund (grant No. 14ZDB153).

## References

- [Atmanspacher, 2012] H. Atmanspacher. Dual-aspect monism à la Pauli and Jung. *Journal of Consciousness Studies*, 19(9-10):96–120, 2012.
- [Busemeyer and Bruza, 2012] J. R. Busemeyer and P. D. Bruza. *Quantum models of cognition and decision*. Cambridge University Press, 2012.
- [Chow, 1998] T. Y. Chow. The surprise examination or unexpected hanging paradox. *American Mathematical Monthly*, pages 41–51, 1998.
- [Gleason, 1957] A. M. Gleason. Measures on the closed subspaces of a hilbert space. *Journal of Mathematics and Mechanics*, 6(6):885–893, 1957.
- [Gödel, 1931] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38(1):173–198, 1931.
- [Heisenberg and Northrop, 1958] W. Heisenberg and F. S. C. Northrop. *Physics and philosophy: The revolution in modern science*. New York: Harper, 1958.
- [Hou and Song, 2009] Y. Hou and D. Song. Characterizing pure high-order entanglements in lexical semantic spaces via information geometry. In *QI*, pages 237–250. Springer, 2009.
- [Hou et al., 2013] Y. Hou, X. Zhao, D. Song, and W. Li. Mining pure high-order word associations via information geometry for information retrieval. *TOIS*, 31(3):12:1–12:32, 2013.
- [Kritchman and Raz, 2010] S. Kritchman and R. Raz. The surprise examination paradox and the second incompleteness theorem. *Notices of the AMS*, 57(11):1454–1458, 2010.
- [Lvovsky, 2004] A. I. Lvovsky. Iterative maximum-likelihood reconstruction in quantum homodyne tomography. *J. Opt. B*, 6(6):S556–S559, 2004.
- [Melucci and van Rijsbergen, 2011] M. Melucci and K. van Rijsbergen. Quantum mechanics and information retrieval. In *Advanced Topics in Information Retrieval*, pages 125–155. Springer, 2011.
- [Metzler and Croft, 2005] D. Metzler and W. B. Croft. A markov random field model for term dependencies. In *SIGIR*, pages 472–479. ACM, 2005.
- [Nakahara and Amari, 2002] H. Nakahara and S. Amari. Information-geometric measure for neural spikes. *Neural Computation*, 14(10):2269–2316, 2002.
- [Nielsen and Chuang, 2010] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2010.
- [Piwowarski et al., 2010] B. Piwowarski, I. Frommholz, M. Lalmas, and K. van Rijsbergen. What can quantum theory bring to information retrieval. In *CIKM*, pages 59–68. ACM, 2010.
- [Rieffel and Polak, 2000] E. Rieffel and W. Polak. An introduction to quantum computing for non-physicists. *ACM Computing Surveys*, 32(3):300–335, 2000.
- [Song and Croft, 1999] F. Song and W. B. Croft. A general language model for information retrieval. In *CIKM*, pages 316–321. ACM, 1999.
- [Song et al., 2010] D. Song, M. Lalmas, C. J. van Rijsbergen, and et al. How quantum theory is developing the field of information retrieval. In *AAAI Fall Symposium*, pages 105–108, 2010.
- [Sordoni et al., 2013] A. Sordoni, J.-Y. Nie, and Y. Bengio. Modeling term dependencies with quantum language models for IR. In *SIGIR*, pages 653–662. ACM, 2013.
- [Sordoni et al., 2014] A. Sordoni, Y. Bengio, and J.-Y. Nie. Learning concept embeddings for query expansion by quantum entropy minimization. In *AAAI*, pages 1586–1592, 2014.
- [van Rijsbergen, 2004] C. J. van Rijsbergen. *The geometry of information retrieval*. Cambridge University Press, 2004.
- [Wittgenstein, 1953] L. Wittgenstein. Philosophical investigations. *London, Basic Blackw*, 1953.
- [Wittgenstein, 1961] L. Wittgenstein. *Tractatus logico-philosophicus, the German text of Logisch-philosophische Abhandlung, with a new translation by DF Pears and BF McGuinness, and with the introd. by Bertrand Russell*. London, Routledge, 1961.
- [Zhai, 2008] C. Zhai. Statistical language models for information retrieval a critical review. *Foundations and Trends in Information Retrieval*, 2(3):137–213, 2008.
- [Zhao et al., 2011] X. Zhao, P. Zhang, D. Song, and Y. Hou. A novel re-ranking approach inspired by quantum measurement. In *ECIR*, pages 721–724. Springer, 2011.
- [Zuccon and Azzopardi, 2010] G. Zuccon and L. Azzopardi. Using the quantum probability ranking principle to rank interdependent documents. In *ECIR*, pages 357–369. Springer, 2010.