

Automata Modeling for Cognitive Interference in Users' Relevance Judgment

Peng Zhang¹, Dawei Song¹, Yuexian Hou^{1,2}, Jun Wang¹, Peter Bruza³

¹School of Computing, The Robert Gordon University, UK.

²School of Computer Sci. & Tec., Tianjin University, China.

³Queensland University of Technology, Australia

Email: p.zhang1@rgu.ac.uk

Abstract

Quantum theory has recently been employed to further advance the theory of information retrieval (IR). A challenging research topic is to investigate the so called quantum-like interference in users' relevance judgment process, where users are involved to judge the relevance degree of each document with respect to a given query. In this process, users' relevance judgment for the current document is often interfered by the judgment for previous documents, due to the interference on users' cognitive status. Research from cognitive science has demonstrated some initial evidence of quantum-like cognitive interference in human decision making, which underpins the user's relevance judgment process. This motivates us to model such cognitive interference in the relevance judgment process, which in our belief will lead to a better modeling and explanation of user behaviors in relevance judgement process for IR and eventually lead to more user-centric IR models. In this paper, we propose to use probabilistic automaton (PA) and quantum finite automaton (QFA), which are suitable to represent the transition of user judgment states, to dynamically model the cognitive interference when the user is judging a list of documents.

Introduction

Increasingly, information retrieval (IR) techniques are underpinning daily information seeking and management tools, such as the popular search engines. Despite the success of IR applications, there has been an urging demand for advanced theoretical frameworks, especially to support the emerging context-sensitive and user-centric IR systems which are compatible with human information processing and cognition. Following van Rijsbergen's pioneering work (2004), which has shown the potential of quantum theory (QT) to subsume the major IR models into a single mathematical formalism in Hilbert vector spaces, researchers proposed many QT-based models in a number of areas, e.g., contextual IR (Melucci 2008), interactive IR (Piwowski and Lalmas 2009a, 2009b), lexical semantic spaces (Bruza and Cole 2005; Hou and Song 2009), and quantum-like interference in IR (Zuccon et al. 2009, 2010; Melucci 2010).

This paper is a step further along the direction of investigating the quantum-like interference in a central IR process - user's relevance judgment, where users are involved to judge the relevance degree of documents with respect to a given query. In this process, users' relevance judgment for the current document is often interfered by the judgment for previously seen documents, as a result of cognitive interference. A typical scenario is that, after finding a more relevant document, a user may lower the relevance degree of a previously judged document. For example, when a user judges whether a document d_0 about "the theory of relativity" is relevant to the query "Albert Einstein", the user might initially think it is 90% relevant. However, after viewing another document d_1 introducing all the aspects of Einstein, the user is likely to consider d_1 as almost 100% relevant, yet the relevance of d_0 may be lowered accordingly. Capturing such cognitive interference, in our belief, will lead to more user-centric IR models.

Research from the cognitive science (Khrennikov 2004; Busemeyer et al. 2007; 2009) has demonstrated initial evidence about the quantum (or quantum-like) nature of the cognitive interference involved in human decision making, which underpins the relevance judgment process in IR. The cognitive experiment showed that the participants' probability for making certain decision could be interfered by a previous categorization task in a non-classical manner: the classical (Kolmogorov) law of total probability (CLTP) was violated. It has been argued that classical Markov model with transition probabilities from category states to decision states was insufficient to model the probability differences, while the quantum model provided a natural way to bridge the gap between non-interfered probabilities and interfered probabilities.

This sheds light for us on the formal modeling of the cognitive interference in the users' relevance judgement process. However, different from the approach in cognitive science (Busemeyer et al. 2009), we will present how an alternative Markov approach, in which transitions are between decision states, rather than (transitions) from category states to decision states, is able to model the above cognitive interference. Similarly, the transition between users' judgment states can model the corresponding cognitive interference in the relevance judgment process. The judgment states correspond to the relevance degrees of documents, e.g., *relevant*

(*R*), *partially relevant* (*P*), and *irrelevant* (*I*).

Formally, we propose to use a probabilistic automaton (PA) (Tzeng 1992), which is a generalization of Markov models, to dynamically model the cognitive interference when user is judging a list of documents. Specifically, the states in PA represent the judgment states of users; different input symbols of PA represent different kinds of cognitive interference generated by viewing other documents; and each symbol is associated with a (Markov) transition matrix to encode the transition probabilities among judgment states. Moreover, PA can accept a finite string of symbols, making it possible to dynamically model the cognitive interference when users are judging a ranking of documents.

In addition to PA, we also present its quantum counterpart, namely quantum finite automaton (QFA) (Kondacs and Watrous 1997), to model the cognitive interference and show its potential advantages over the PA.

To the best of our knowledge, this is the first attempt to formally model the cognitive interference in the document relevance judgement process. In the IR literature, the dynamic nature of relevance and the effects of document presentation order on relevance judgment have been addressed in (Mizzaro 1996; Eisenberg and Barry 1988). Recently, Zuccon et al. (2009; 2010) proposed a quantum probability ranking principle (QPRP), which incorporates the approximation (i.e., the interference term estimated by the similarity between non-ranked document and ranked documents) of the quantum interference term into the classical probability ranking principle (PRP) to revise the document ranking scores. QPRP is the first successful application of quantum-like IR methods and has achieved a promising performance on large scale collections. However, QPRP can be regarded as a system-centric approach, since it is concerned about the document ranking by the system. Different from QPRP, our approach is focused on user's relevance judgment process: having judged the relevance of other documents will lead to the user's change of cognitive status and in turn interfere with his/her judgment of current document. In our opinion, system-centric approach implicitly assumes that the relevance is a reality feature (or an objective property) of documents, while the user-centric approach is to investigate the non-reality feature (or subjective property) of relevance, which coincides with the non-reality principle of quantum theory.

Cognitive Interference in Human Decision Making

In this section, we start with a description of Busemeyer et al.'s cognitive experiment in human decision making and their formulation on the cognitive interference. Then, we present an alternative Markov approach to model the cognitive interference.

In (Busemeyer and Wang 2007; Busemeyer, Wang, and Lambert-Mogiliansky 2009), researchers presented some experimental results that demonstrated the nonclassical nature of the cognitive interference, evidenced by the violation of the classical law of total probability (CLTP), in a human decision making task. Participants were shown pictures of

human faces and asked to make a decision about how they would react to the faces. Two experimental settings were predefined. In setting 1, the participants were asked to decide (D-only) whether to take a 'attack (A)' or 'withdraw (W)' action. In setting 2, the participants were asked to first categorize (C step) the face as belonging to either a 'good (G)' or a 'bad (B)' guy, and then decide (C-then-D) their actions (A or W). The experiment was carried out in a well controlled environment to ensure less distraction to the participants and more randomness in assigning participants to different experimental settings.

Violation of CLTP in Human Decision

Through this experiment, various probabilities¹, including $P(A)$ and $P(W)$ in setting 1, and $P(G)$, $P(A|G)$, $P(W|G)$, $P(B)$, $P(A|B)$ and $P(W|B)$ in setting 2, can be obtained. These probabilities are denoted for short as p_A and p_W in setting 1, and p_G , T_{AG} , T_{WG} , p_B , T_{AB} and T_{WB} in setting 2, respectively. The following observation from the experiment indicates the violation of CLTP:

$$\begin{bmatrix} p_A \\ p_W \end{bmatrix}^T \neq \begin{bmatrix} p_G \cdot T_{AG} + p_B \cdot T_{AB} \\ p_G \cdot T_{WG} + p_B \cdot T_{WB} \end{bmatrix}^T \quad (1)$$

The above formula shows that the non-interfered probabilities on the left-hand side are different from the interfered probabilities on the right hand side.

A Markov Model Explanation

To model and explain the phenomenon, Busemeyer et al. (2009) considered the right-hand side of Formula 1 as a Markov model. They defined two category states, G and B, and two decision states, A and W. The 2×1 row vector $P_C = [p_G \ p_B]$ represents the initial probabilities of the two category states in setting 2. The 2×2 matrix

$$T = \begin{bmatrix} T_{AG} & T_{WG} \\ T_{AB} & T_{WB} \end{bmatrix} \quad (2)$$

can represent the probabilities of transiting from each category states to each decision states. Specifically, $T_{j,i}$ represents probability of transiting from category state j (G or B) to decision state i (A or W). T is transition matrix (a right stochastic matrix), which means each row of T sums to one. Then, the right-hand side of Formula 1 equals to $P_C \cdot T$, a standard Markov process. Let $P_F = [p_{FA} \ p_{FW}] = P_C \cdot T$, where p_{FA} and p_{FW} are the probabilities predicted by the Markov Model. Hence, Formula 1 can be rewritten as:

$$\begin{bmatrix} p_A \\ p_W \end{bmatrix}^T \neq \begin{bmatrix} p_{FA} \\ p_{FW} \end{bmatrix}^T = \begin{bmatrix} p_G \cdot T_{AG} + p_B \cdot T_{AB} \\ p_G \cdot T_{WG} + p_B \cdot T_{WB} \end{bmatrix}^T \quad (3)$$

It turns out that the predicted probabilities of the Markov model are still different from the probabilities in the setting 1. Moreover, this Markov approach still can not bridge the probability differences in the two different settings.

¹In this section, for a better readability, some notations and formulation in Sections 2.1 and 2.2 of (Busemeyer, Wang, and Lambert-Mogiliansky 2009) are slightly modified

Quantum Model Explanation

Later, Busemeyer et al. (2009) proposed to use quantum theory to explain the violation of CLTP. In quantum theory, the probabilities are derived from the squared magnitudes of the amplitudes. For example, an amplitude q is a complex number $q = r \cdot [\cos \theta + i \sin \theta]$, where $i = \sqrt{-1}$, $r \in [0, 1]$ is the magnitude, $\theta \in [0, 2\pi)$ is the phase. The squared magnitude $q \cdot q^* = r^2 \leq 1$.

Two category states, $|G\rangle$ and $|B\rangle$, as well as two decision states, $|A\rangle$ and $|W\rangle$, were defined. The 2×1 row vector $Q_C = [q_G \ q_B]$ represents the initial amplitudes for the category states, where $|q_G|^2 = p_G$ and $|q_B|^2 = p_B$. The 2×2 matrix

$$U = \begin{bmatrix} U_{AG} & U_{WG} \\ U_{AB} & U_{WB} \end{bmatrix} \quad (4)$$

represents the amplitudes for transiting from each category state to each decision state, where U is a unitary matrix² and $|U_{i,j}|^2 = T_{i,j}$. Then, the predicted amplitudes by the quantum model can be represented as

$$Q_F = \begin{bmatrix} Q_{FA} \\ Q_{FW} \end{bmatrix}^T = Q_C \cdot U = \begin{bmatrix} q_G \cdot U_{AG} + q_B \cdot U_{AB} \\ q_G \cdot U_{WG} + q_B \cdot U_{WB} \end{bmatrix}^T \quad (5)$$

The corresponding probabilities by quantum model are

$$\begin{aligned} & \begin{bmatrix} |Q_{FA}|^2 \\ |Q_{FW}|^2 \end{bmatrix}^T = \\ & \begin{bmatrix} |q_G|^2 |U_{AG}|^2 + |q_B|^2 |U_{AB}|^2 + 2 \cdot |q_G| |U_{AG}| |q_B| |U_{AB}| \cos(\theta) \\ |q_G|^2 |U_{WG}|^2 + |q_B|^2 |U_{WB}|^2 + 2 \cdot |q_G| |U_{WG}| |q_B| |U_{WB}| \cos(\tau) \end{bmatrix}^T \\ & \approx \begin{bmatrix} p_G \cdot T_{AG} + p_B \cdot T_{AB} + \text{Interference}(\theta) \\ p_G \cdot T_{WG} + p_B \cdot T_{WB} + \text{Interference}(\tau) \end{bmatrix}^T \end{aligned} \quad (6)$$

It shows that the quantum probabilities do not necessarily obey the CLTP due to the existence of the interference term, e.g., $2 \cdot |q_G| |U_{AG}| |q_B| |U_{AB}| \cos(\theta)$. As a result, these probabilities $[|Q_{FA}|^2 \ |Q_{FW}|^2]$ of quantum model are different from the predicted probabilities $[p_{FA} \ p_{FW}]$ of the Markov model in Formula 3. In (Busemeyer, Wang, and Lambert-Mogiliansky 2009), they have shown that using a well trained $\cos(\theta)$, $\cos(\tau)$ and fitted unitary matrices, $[|Q_{FA}|^2 \ |Q_{FW}|^2]$ can predict $[p_A \ p_W]$ more exactly.

An Alternative Markov Approach

We think that the cognitive interference observed in the above human decision task can be modeled by an alternative Markov approach.

First, we assume in general that if there does not exist cognitive interference, the CLTP should hold; otherwise, the CLTP can be violated. For the experiment in (Busemeyer, Wang, and Lambert-Mogiliansky 2009), this assumption means that if there was no cognitive interference in setting 2 (i.e., the participants' C-step does not interfere with the D-step), the CLTP should hold. Otherwise, the probability difference between both sides can be generated.

²To make sure that U is a unitary matrix, $|U_{i,j}|^2$ just need to be fitted to approximate $T_{i,j}$

If no cognitive interference exists in setting 2, we let T_{AG}^N , T_{AB}^N , T_{WG}^N and T_{WB}^N be the non-interfered conditional probabilities in setting 2, which correspond to the interfered ones T_{AG} , T_{AB} , T_{WG} and T_{WB} in Formula 1, respectively. Notice that different from $T(\cdot)$, the $T(\cdot)^N$ are just for theoretical analysis and are actually not observed from the experimental data. We let

$$P_N = \begin{bmatrix} p_{NA} \\ p_{NW} \end{bmatrix}^T = \begin{bmatrix} p_G \cdot T_{AG}^N + p_B \cdot T_{AB}^N \\ p_G \cdot T_{WG}^N + p_B \cdot T_{WB}^N \end{bmatrix}^T \quad (7)$$

where p_{NA} and p_{NW} denote the non-interfered probabilities with respect to action A and W, respectively in setting 2. According to our assumption, the CLTP should hold, which means

$$\begin{bmatrix} p_A \\ p_W \end{bmatrix}^T = \begin{bmatrix} p_{NA} \\ p_{NW} \end{bmatrix}^T \quad (8)$$

where the left-hand side is actually the left-hand side of Formula 1, while the right-hand side is just for theoretical analysis but not observed from the data.

If there exists cognitive interference in setting 2, the CLTP does not hold, which gives us

$$P_I = \begin{bmatrix} p_{IA} \\ p_{IW} \end{bmatrix}^T = \begin{bmatrix} p_G \cdot T_{AG} + p_B \cdot T_{AB} \\ p_G \cdot T_{WG} + p_B \cdot T_{WB} \end{bmatrix}^T \quad (9)$$

where the $P_I = [p_{IA} \ p_{IW}]$ represents the interfered probabilities in setting 2, and are actually equivalent to the right-hand side of Formula 1. Hence, the violation of CLTP in Formula 1 can be rewritten as:

$$\begin{bmatrix} p_A \\ p_W \end{bmatrix}^T \neq \begin{bmatrix} p_{IA} \\ p_{IW} \end{bmatrix}^T \quad (10)$$

From the above four Formulae, it turns out that the problem to explain the violation of the CLTP in Formula 10 is equivalent to the problem of explaining the difference between P_N and P_I (i.e., two right-hand sides of Formula 8 and 10, respectively) and how to bridge the difference. Notice that under our assumption, the quantum explanation in Formula 6 can also be regarded as an approach to explain the difference between P_N and P_I and use an interference term to bridge the difference.

The difference between P_N and P_I is generated by the difference between $T(\cdot)^N$ and $T(\cdot)$ (see Formula 7 and Formula 9). We think this is due to the transition between the decision states (A or W), which can be formulated by a standard Markov transition matrix (or called a right stochastic matrix):

$$M(\alpha) = \begin{bmatrix} M_{AA} & M_{AW} \\ M_{WA} & M_{WW} \end{bmatrix} \quad (11)$$

where $M(\alpha)_{i,j}$ ($i, j \in \{A, W\}$) represents the probabilities of transiting from a decision state i to another decision state j , and different $M(\alpha)$ ³ represent different cognitive interference effects of the category state α ($\alpha \in \{G, B\}$). Hence,

³It should be noted that $M(\alpha)$ is different from T (see Formula 2), in the sense that $M(\alpha)$ represents the transition probabilities between decision states, while T represents the transition probabilities from category states to decision states.

the cognitive interference process can be formulated as follows:

$$\begin{aligned} [T_{AG} \ T_{WG}] &= [T_{AG}^N \ T_{WG}^N] \times M(G) \\ [T_{AB} \ T_{WB}] &= [T_{AB}^N \ T_{WB}^N] \times M(B) \end{aligned} \quad (12)$$

If each $M(i)$ is an identity matrix, it means that categorizing the face does not interfere with participants' decisions on their actions. Otherwise, according to Formula 12, the C-step interferes with the D-step, making the difference between $T_{(\cdot)}^N$ and $T_{(\cdot)}$, and then the difference between P_N and P_I . The transition matrices $M(G)$ and $M(B)$ can be regarded as a connection between $T_{(\cdot)}^N$ and $T_{(\cdot)}$, and then bridge the difference between P_N and P_I .

For example, in Formula 12, $[T_{AG}^N \ T_{WG}^N] = [0.8 \ 0.2]$ means that without interference in setting 2, for the participants who categorized the face as belonging to a 'good' (G) guy, 80% of them decided to 'attack' (A) and 20% decided to 'withdraw' (W). If the transition matrix $M(G)$ is an identity matrix, then $[T_{AG} \ T_{WG}] = [T_{AG}^N \ T_{WG}^N]$, which indicates that the category state G does not interfere with the decision states. Otherwise, suppose the transition matrix $M(G) = \begin{bmatrix} 0.4 & 0.6 \\ 0.7 & 0.3 \end{bmatrix}$, which means that for the participants (see the transition probabilities in the first row of $M(G)$) who first decided to attack, 40% would stick to their initial decisions while 60% would change their decisions to withdraw; for the participants (see the second row) who first decided to withdraw, 70% would change their decisions to attack while 30% would stay on their initial decisions. After the cognitive interference represented by $[T_{AG}^N \ T_{WG}^N] \times M(G)$, the interfered probabilities are $[T_{AG} \ T_{WG}] = [0.8 \ 0.2] \times \begin{bmatrix} 0.4 & 0.6 \\ 0.7 & 0.3 \end{bmatrix} = [0.46 \ 0.54]$, which is different from the non-interfered probabilities.

Further Comparison between the Above Models

From Formula 12, one may think that we are using transition matrices and non-interfered probabilities to predict the interfered probabilities. This is different from that in the quantum model (Formula 6), where they are using an interference term and interfered probabilities to predict the non-interfered ones. We should mention that: firstly, through the inverse operation of Formula 12, one can also predict the non-interfered probabilities; secondly, to predict the interfered decisions is actually more important in the user relevance judgment process.

Automata Modeling for Cognitive Interference in Relevance Judgment

In the previous section, we have shown the possibility of using Markov approach to model the cognitive interference in human decision making context.

Now, we investigate the cognitive interference during the user relevance judgment process⁴. Here, the cognitive inter-

⁴We admit that relevance judgment can be affected by a lot of factors, such as dynamic information needs, contents, genre and

ference means that user relevance judgment for some documents often interferes with user judgment for other documents. Our general aim is to dynamically model the cognitive interference while users are scanning a ranking of documents.

The above aim is highly related to interactive IR tasks. For example, if the users have read and explicitly or implicitly judged some of the top-ranking documents, the relevance scores of other documents should be revised accordingly. The score revision process should consider the certain groups/types of users' cognitive interference, which can be learned from the model proposed in this paper.

Basic Idea

Now, we describe the basic idea of modelling the cognitive interference between users judging two document d_0 and d_1 , using the standard Markov way described in the previous section.

Assume there are three relevance degrees, R , P , and I , which represent the relevance degrees *relevant*, *partially relevant* and *irrelevant*, respectively, of a document with respect to the given query q . We let relevance degrees be judgment states of users, which correspond to the decision states in the previous section. Let a 1×3 row vector $\mathbf{w} = [p_R, p_P, p_I]$ be the judgment states distribution (states distribution for short) to represent the judgment result. Specifically, p_R , p_P and p_I represent the percentage of users who judge the document as relevant (R), partially relevant (P), irrelevant (I). Let \mathbf{w}_0 be non-interfered states distribution of document d_0 , and $\mathbf{w}_0^{(\alpha)}$ be interfered states distribution of d_0 . The cognitive interference is generated by judging another document d_1 , and the interference is encoded in a transition matrix

$$M(\alpha) = \begin{bmatrix} M_{RR} & M_{RP} & M_{RI} \\ M_{PR} & M_{PP} & M_{PI} \\ M_{IR} & M_{IP} & M_{II} \end{bmatrix} \quad (13)$$

where $M_{i,j}$ ($i, j \in \{R, P, I\}$) represents the probabilities of transiting from judgment state i to judgment state j . Different interference by document d_1 has different $M(\alpha)$, which will be discussed later.

This $M(\alpha)$, which represents the cognitive interference among judgment states, borrows the same idea from the M in Formula 11. Furthermore, $M(\alpha)$ can bridge the gap between the non-interfered states distribution \mathbf{w}_0 and interfered counterpart $\mathbf{w}_0^{(\alpha)}$ by $\mathbf{w}_0^{(\alpha)} = \mathbf{w}_0 \times M(\alpha)$ ⁵. This is similar to the idea that $M(G)$ and $M(B)$ can connect the difference between P_N and P_F .

In the next section, we will adopt the probabilistic automata (PA), which is a generalization of Markov models,

even the retrieval interface. Here, we pay more attention to the change of user relevance judgment results of documents with respect to a given query. Actually, we also studied the cognitive interference in the relevance judgment of the potential queries (indicating user information needs) with respect to a given document, which is out of the scope of this paper.

⁵Here, in order to keep the formulation simpler, we do not include the conditional probabilities.

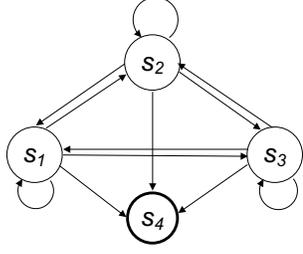


Figure 1: An example for Probabilistic Automaton (PA). In this PA, the states set is $S = \{s_1, s_2, s_3, s_4\}$, where s_4 is the final state. Each PA has an alphabet Σ . For each symbol $\alpha \in \Sigma$, $M(\alpha)$ is the transition matrix to represent the probabilities of the transition among states.

to model the cognitive interference in the relevance judgment. The reason why we adopt PA is that: firstly, PA offers different input symbols, which can represent different kinds of cognitive interference generated by viewing other documents; secondly, PA can accept a finite string of symbols, making it possible and natural to dynamically model the cognitive interference when users are judging a ranking of documents.

In addition, the cognitive interference modeled by quantum finite automaton (QFA) will also be presented. Here, we clarify that cognitive interference using the classical Markov model, e.g. PA, may only be able to explain the violation of CLTP described by Formula 1 or other similar phenomenon. In the following, we will pay more attention to the PA, since we think it is easier to implement in IR and it is powerful enough for our task. On the other hand, we will further show the potential advantages of QFA, which is more general and powerful.

PA Modeling for Cognitive Interference

We propose to use probabilistic automaton (PA), which is a generalization of the concept of Markov model, to model the cognitive interference. We are more focused on the model rather than examples or observations, since the difference between non-interfered judgment and interfered judgment is a widely-accepted fact in IR but less attention has been paid to formal modeling of it. We first give the formal definition of probabilistic automaton (PA).

Definition of PA Before giving the Definition 1 of PA, we should mention that a (row) vector is *stochastic* if all its entries are greater than or equal to 0 and sum to 1. Accordingly, a matrix is *stochastic* if all its row vectors are stochastic. We let $\mathcal{M}_{(m,n)}$ be the set of all $m \times n$ stochastic matrices.

Definition 1 A probabilistic automaton (PA) U is a 5-tuple $(S, \Sigma, M, \mathbf{w}, F)$, where $S = \{s_1, \dots, s_m\}$ is a finite set of states, Σ is an input alphabet, M is a function from Σ into $\mathcal{M}_{(m,m)}$, \mathbf{w} is a (m) -dimensional stochastic row vector, and $F \subseteq S$ is a set of final states.

The vector \mathbf{w} denotes the initial states distribution whose k^{th} component is the initial probability of the state s_k ($s_k \in S$). For each symbol $\alpha \in \Sigma$, $M(\alpha)$ is a $m \times m$ stochastic

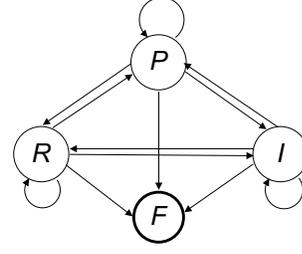


Figure 2: Illustration of cognitive interference modeling by PA. The states distribution of PA can represent users relevance judgment. The transition matrices of PA can be used to model the cognitive interference when users are judging documents.

matrix. The value $M(\alpha)_{k,t} = \Pr(s_k \xrightarrow{\alpha} s_t)$ is the probability that the state s_k transits to s_t after reading α . For the PA, after reading α , the new states distribution becomes $\mathbf{w}M(\alpha)$. If then another symbol $\beta \in \Sigma$ is read in, the next states distribution of PA will be $\mathbf{w}M(\alpha)M(\beta)$. Therefore, we can extend the domain of function M from Σ to Σ^* in a standard way, i.e., $M(\alpha\beta) = M(\alpha)M(\beta)$, where $\alpha\beta$ is a string of symbols.

An indicative example of a 4-state PA is given in Figure 1. Suppose the initial distribution $\mathbf{w} = [0.5, 0.3, 0.2, 0]$,

$$\text{and } M(\alpha) = \begin{bmatrix} 0.6 & 0.35 & 0.05 & 0 \\ 0.1 & 0.9 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \text{ After reading}$$

α , the states distribution of this PA will be $\mathbf{w}M(\alpha) = [0.330, 0.445, 0.225, 0]$. The values 0.6, 0.35, 0.05 and 0 in the first row of this $M(\alpha)$ are the probabilities that s_1 transits to s_1, s_2, s_3 and s_4 , respectively. In general, the k^{th} row in $M(\alpha)$ represents the probabilities that the state s_k transits to every s_t ($s_t \in S$).

PA Representation for Users Judgment Results The states s_1, \dots, s_m and their distribution in PA can represent the uses relevance judgment of a document. Recall the example illustrated in the ‘‘Basic Idea’’ section. Now, we are going to reuse the same example in the PA framework. As shown in Figure 2, the PA states $\{s_1, s_2, s_3, s_4\} = \{R, P, I, F\}$. R, P , and I represent the relevance degrees (also called judgment states), *relevant*, *partially relevant* and *irrelevant*, respectively, of a document with respect to the given query q . The state F denotes that the user quits or finishes the judging process. For the states distribution \mathbf{w} , the probability p_1, p_2, p_3 or p_4 corresponds to the percentage of users who judge the document as relevant (R), partially relevant (P), irrelevant (I), or finish the judgment (F), respectively.

In general, PA can represent a finite number of relevance degrees. Let the first $m-1$ PA states s_1, \dots, s_{m-1} represent the relevance degrees which are monotonically decreasing. That means s_1 denotes *totally relevant* and s_{m-1} denotes *absolutely irrelevant*. The final state s_m indicates that the user quits or finishes the relevance judgment. Then, we let

$$\mathbf{w}_i = [p_{i,1}, \dots, p_{i,m}] \quad (14)$$

represent the *initial* relevance judgments of users with re-

spect to d_i , where $p_{i,k}$ ($k = 1, \dots, m$) denotes the probability (e.g., the percentage of users) of judging d_i as relevance degree s_k . Note that the *initial* judgment means that the users only judge document d_i , i.e., without interference.

Cognitive Interference Modeling when Judging two Documents Now, we describe how this PA can be used to model the process that judging a document d_0 is interfered by judging another document d_1 . We first map the initial relevance judgment result to the ranking score, in order to quantify the topical relevance of documents. Specifically, this mapping is as follows:

$$\psi(d_i) = \mathbf{w}_i \times \Lambda \quad (15)$$

where $\psi(d_i)$ is the initial relevance score of d_i , and Λ is a column vector $[m, m-1, \dots, 1]^T$. Then, $\psi(d_i) = \sum_{k=1}^m p_{i,k} \times (m-k+1)$. This is to ensure that the more users judged higher relevance degrees, the higher the overall relevance score would be⁶.

Recall that PA has several input symbols in the alphabet Σ and different symbols can represent different kinds of interference. Specifically, we define three input symbols, a , b and c , which correspond to the following three conditions, i.e., 1) $\psi(d_1) > \psi(d_0)$, 2) $\psi(d_1) < \psi(d_0)$ and 3) $\psi(d_1) = \psi(d_0)$. Obviously, these three conditions reflect the relative relevance of two documents d_1 and d_0 . Let another input symbol e denote that the user will finish the relevance judgment task. Formally, the input symbol generated by judging document d_1 is formulated as:

$$\alpha(d_1) = \begin{cases} a, & \psi(d_1) > \psi(d_0) \\ b, & \psi(d_1) < \psi(d_0) \\ c, & \psi(d_1) = \psi(d_0) \\ e, & \text{Judgment finished} \end{cases} \quad (16)$$

To model the process that judging d_0 is interfered by judging d_1 , the initial states distribution of PA is set to \mathbf{w}_0 and the input symbol is generated by judging d_1 . For each symbol, a transition matrix $M(\alpha)$ will be learned from user judgment data and reflect the states transition probabilities. These probabilities actually reflect users cognitive interference. For example, $M(\alpha)_{k,t}$ denote that the user will change their relevance judgment from relevance degree q_k to degree q_t , after reading another document that generates an input factor α . The whole interference process can be formulated as

$$\mathbf{w}_0^{(\alpha)} = \mathbf{w}_0 \times M(\alpha) \quad (17)$$

where $\mathbf{w}_0^{(\alpha)}$ is the *interfered* relevance judgment of d_0 and $\alpha = \alpha(d_1)$ is generated by judging d_1 .

Cognitive Interference Modeling when Judging a Ranking of Documents Generally, our aim is to dynamically model the cognitive interference of users when they are judging a list of documents $d_1 d_2 \dots d_n$. This means for each j ($1 \leq j \leq n$), PA is expected to model how the judgment for $d_1 d_2 \dots d_j$ interferes with the judgment for document d_i ($1 \leq i \leq n$).

⁶There may be other choice for Λ in Equation 15, which is left as our future work.

When we investigate the interference⁷ on \mathbf{w}_i (*w.r.t.* d_i) imposed by \mathbf{w}_j (*w.r.t.* d_j), the initial states distribution of PA is set as \mathbf{w}_i , and the input symbol is generated by judging d_j using the following formula:

$$\alpha(d_j) = \begin{cases} a, & \psi(d_j) > \psi(d_i) \\ b, & \psi(d_j) < \psi(d_i) \\ c, & \psi(d_j) = \psi(d_i) \\ e, & \text{Judgment finished} \end{cases} \quad (18)$$

To model the process that having judged $d_1 d_2 \dots d_j$ interferes with judging d_i ($1 \leq i \leq n$), we extend the domain of the function M from Σ to Σ^* , as described in first paragraph below Definition 1. After judging a ranking of documents $d_1 d_2 \dots d_j$, a string $x = \alpha(d_1)\alpha(d_2)\dots\alpha(d_j)$ is generated. Accordingly, we have $M(x) = M(\alpha(d_1)\alpha(d_2)\dots\alpha(d_j)) = \prod_{k=1}^j M(\alpha(d_k))$. Then, the interfered states distribution $\mathbf{w}_i^{(x)}$ of d_i imposed by the judgment of $d_1 d_2 \dots d_j$ can be obtained by

$$\mathbf{w}_i^{(x)} = \mathbf{w}_i \times M(x). \quad (19)$$

Note that the above formulation considers the presentation order of documents. Different document order generates different string x of symbols, leading to different transition matrix $M(x)$ and $\mathbf{w}_i^{(x)}$.

Comparing PA with other Classical Models Probabilistic automaton (PA) is a non-deterministic finite automaton (NFA). In this section, we will justify why we choose PA, instead of deterministic finite automaton (DFA), or other graphical models (e.g., hidden Markov model (HMM) or Bayesian network (BN)), to model the cognitive interference.

Firstly, we compare PA with the deterministic finite automaton (DFA). The definition (Tzeng 1992) of DFA is as follows:

Definition 2 A *deterministic finite automaton (DFA)* A is a 5-tuple $(S, \Sigma, \tau, s_1, F)$, where S is a finite set of states, Σ is an input alphabet, τ is a function from $S \times \Sigma$ into S , s_1 is the initial state, and $F \subseteq S$ is a set of final states.

Recall that in PA, each state can have several possible next states after reading a symbol α . From the DFA Definition 2, however, we can see that after reading a symbol, for every state in DFA, the next possible state is uniquely determined. This means that after judging another document, users will change their judgment from one relevance degree (e.g., absolutely relevant) to another unique relevance degree (e.g., absolutely irrelevant). Therefore, DFA cannot model the probabilistic transition from one relevance degree to many possible relevance degrees, which is necessary if the cognitive interference modeling consider a large number of users with different behavioral trends.

Secondly, compared to other graphical models, such as Hidden Markov Model (HMM) and Bayesian Network (BN), PA is also more suitable for our task. Both PA and

⁷Here, we say \mathbf{w}_j interferes with \mathbf{w}_i , since we think the cognitive interference is on users relevance judgments.

HMM are probabilistic finite-state machines (Vidal et al. 2005) and some link between them has been found (Dupont, Denis, and Esposito 2005)(Vidal et al. 2005). However, it is not straightforward to define the hidden state in our problem. Classical Bayesian network (BN) cannot model the sequential task. Dynamic Bayesian network (DBN) (Ghahramani 1998) can model sequences of variables, but it could be too complex to be not directly applied to our problem. Nonetheless, we would like to investigate how to use DBN for our problem in the future.

QFA Modeling for Cognitive Interference

Quantum finite automata (QFA) are a quantum analog of probabilistic automata. In this section, we describe the main idea of QFA modeling of the cognitive interference, by providing the quantum version of the components in the PA modeling.

Using QFA, the states are denoted as $|s_1\rangle, |s_2\rangle, \dots, |s_m\rangle$, where $|s_k\rangle$ ($1 \leq k \leq m$) denotes the k^{th} judgment states. The states distribution for the document d_i is denoted as $|w_i\rangle = [q_{i,1}, q_{i,2}, \dots, q_{i,m}]$, where $q_{i,k}$ denotes the amplitude of the k^{th} judgment state, and $|q_{i,k}|^2 = p_{i,k}$ in Formula 14. We adopt the same mapping function $\psi(d_i)$ in Formula 15 and the interference symbol generation function in Formula 18 for QFA modeling. For each input symbol α , the corresponding quantum transition matrix is a unitary matrix $U(\alpha)$. We denote the interfered states distribution as $|w_i^\alpha\rangle$, where

$$|w_i^\alpha\rangle = |w_i\rangle U(\alpha) \quad (20)$$

After judging a ranking of documents that generate a sequence of interference input symbol, denoted as $x = \alpha(d_1)\alpha(d_2) \dots \alpha(d_j)$, the interfered states distribution for document d_i is:

$$|w_i^x\rangle = |w_i\rangle U(x) \quad (21)$$

where $U(x) = \prod_{k=1}^j U(\alpha(d_k))$. Note that the above quantum Markov process is different to that proposed by (Busemeyer, Wang, and Lambert-Mogiliansky 2009), in the sense that the former is the transition of judgment states, while the latter is the transition from category to decision states.

The above quantum formulation has its advantage in that the quantum probabilities naturally do not need to obey classical probability law, e.g. CLTP. On the other hand, PA may just be able to explain the violation of CLTP in the sense of the difference between the non-interfered (or initial) and interfered states distribution. Moreover, we consider that QFA is stronger than PA in that QFA can naturally offer (potentially) infinite memory and the corresponding processing ability. Note that, being aided by (potentially) infinite external storage and dominated by inductive and abstract ability, the cognitive process of human being can handle, at least to some extent, (potentially) infinite information. Hence, QFA would potentially make a more general and powerful approach to model the cognitive process.

Case Study

We designed and carried out a pilot user study to investigate the cognitive interference when users are judging two

documents. In this study, five queries were selected. They are (1) “Albert Einstein”, (2) “Probability”, (3) “Semantic Web”, (4) “Graphical Model”, and (5) “Quantum Mechanics”. The number of queries is small, since the users are often more careful if the task does not take them too much time. For each query, three Wikipedia pages with different relevance degrees were chosen. Recall that PA can model a finite number of relevance degrees. Here, the relevance degrees are ranged from A to E , where A means “totally relevant” and E means “absolutely irrelevant”, then S is $\{A, B, C, D, E\}$. We fixed one page as d_0 (which is mostly likely to be partially relevant, i.e. $B, C, \text{ or } D$), and d_1 is randomly selected from the other two pages before the user judge d_0 .

The users were divided into 2 distinct groups. For each query, we let users in group 1 judge d_0 only (to obtain w_0), and users in group 2 judge d_1 at first (to obtain w_1) and then d_0 (to obtain $w_0^{(\alpha)}$). The reason why we split users into two groups is to avoid any user reading d_0 twice. Users were required to mark the relevance degree from A to E .

The study was carried out through the Amazon Mechanical Turk, which is an Amazon Web service that enables online user studies to be “crowdsourced” to a large number of users quickly and economically. To gain a better controllability of user experiments, it is important to detect suspicious responses (Kittur, Chi, and Suh 2008). Therefore, in our experiment, we removed the data of users who are considered as not reliable in terms of, e.g., without completing most of the queries and without viewing any document for more than a certain time interval. Eventually, for each group, the data of 30 Mechanical Turk users were collected.

At first, we would like to test whether there exists cognitive interference in group 2. We denote the initial judgment result (i.e., without interference) of d_0 in group 1 as w_0 , and the judgment result of d_0 in group 2 as \tilde{w}_0 to indicate the potential interference. We then use Euclidean distance $d(w_0, \tilde{w}_0)$ to measure the difference between w_0 and \tilde{w}_0 . The difference for each query is summarized in Table 1. We can observe that the difference is obvious for every query and indicates the existence of the cognitive interference. More specifically, we found that users tended to lower their relevance degrees of the current document after they viewed a more relevance document, and vice versa. Generally, the initial relevance judgment w_0 on a document d_0 does not hold anymore, once the interference from judging other documents takes effect, leading to the interfered judgment \tilde{w}_0 .

This has practical implications in scenarios such as: after users have judged some documents, the retrieval system is expected to be able to dynamically re-rank other documents according to the interfered relevance. Hence, it is necessary to derive a more reasonable estimation, rather than w_0 , for the real interfered relevance result \tilde{w}_0 .

It is natural to use $w_0^{(\alpha)}$ by Equation 17 to estimate \tilde{w}_0 . Then, we need to test if $d(w_0^{(\alpha)}, \tilde{w}_0)$ is smaller than $d(w_0, \tilde{w}_0)$, where w_0 is used as a baseline for \tilde{w}_0 . If this is true, we can conclude that $w_0^{(\alpha)}$, computed from the proposed PA approach, is a better estimation of \tilde{w}_0 .

Table 1: The Difference between Initial and Interfered Judgment of d_0

Query	#1	#2	#3	#4	#5
$d(\mathbf{w}_0, \tilde{\mathbf{w}}_0)$	0.3713	0.2062	0.2962	0.2708	0.1499

Table 2: The Difference between (Estimated) Interfered and (Real) Interfered Judgment of d_0

Query	#1	#2	#3	#4	#5
$d(\mathbf{w}_0^{(\alpha)}, \tilde{\mathbf{w}}_0)$	0.2323	0.1404	0.2306	0.1833	0.1266

In order to do this test, we assumed that 75% of users data were available and used them for the training of transition matrices $M(\alpha)$. The other 25% of users data was used to test whether $\mathbf{w}_0^{(\alpha)}$ is a better estimation of $\tilde{\mathbf{w}}_0$. 4-fold cross validation was carried out. For training the transition matrix $M(\alpha)$, we simply calculated probabilities $\Pr\{s_k \xrightarrow{\alpha} s_t\}$, by simulating states transition from users data between two groups. Then, we use the trained $M(\alpha)$ to compute $\mathbf{w}_0^{(\alpha)} = \mathbf{w}_0 M(\alpha)$. The difference between $\mathbf{w}_0^{(\alpha)}$ and $\tilde{\mathbf{w}}_0$ for each query⁸ are in Table 2. It can be observed that $d(\mathbf{w}_0^{(\alpha)}, \tilde{\mathbf{w}}_0)$ is much smaller than $d(\mathbf{w}_0, \tilde{\mathbf{w}}_0)$ almost for every query. This implies the potential usefulness of our model for predicting the interfered judgment of users. To the best of our knowledge, there is yet no other result to compare with. We report our first results on real user data.

Conclusions and Future Work

In this paper, we have proposed an alternative Markov approach to explain the quantum-like cognitive interference (in the form of violation of the classical law of total probability) in human decision making which underpins users' relevance judgment process in information retrieval. Then, we proposed to use probabilistic automaton (PA) and quantum finite automaton (QFA) to dynamically model the cognitive interference of users when they are judging a list of documents. A quantitative case study on the real users relevance judgment data collected through a task-based user study demonstrated the suitability and feasibility of the PA modeling for cognitive interference.

In the future, we will investigate in-depth the difference between PA and QFA for cognitive interference modeling. We also need to consider the exact quantity of relevance score differences, and the statistical dependency among documents in the modeling. Moreover, we will refine our experimental set-up according to our general aim, and recruit more users in the relevance judgment experiment.

Additional Authors

Massimo Melucci (University of Padua, email: melo@dei.unipd.it), and John McCall (The Robert

⁸Note that in the real application, it is not necessary to have one transition matrix for only one query. Instead, it is reasonable to learn a transition matrix for a number of queries

Gordon University, email: jm@comp.rgu.ac.uk)

Acknowledgments

We would like to thank Leszek Kaliciak and anonymous reviewers for their constructive comments. This work is supported in part by the UK's EPSRC grant (No.: EP/F014708/2).

References

- Bruza, P., and Cole, R. J. 2005. Quantum logic of semantic space: An exploratory investigation of context effects in practical reasoning. *We will Show Them! Essays in Honour of Dov Gabbay* 1:339–362, College Publications.
- Busemeyer, J. R., and Wang, Z. 2007. Quantum information processing explanation for interactions between inferences and decisions. In *Proceedings of the AAAI Symposium on Quantum Interaction*. AAAI Press.
- Busemeyer, J. R.; Wang, Z.; and Lambert-Mogiliansky, A. 2009. Empirical comparison of markov and quantum models of decision making. *Journal of Mathematical Psychology* 53:423–433.
- Dupont, P.; Denis, F.; and Esposito, Y. 2005. Links between probabilistic automata and hidden markov models: probability distributions, learning models and induction algorithms. *Pattern Recognition* 38(9):1349–1371.
- Eisenberg, M., and Barry, C. 1988. Order effects: A study of the possible influence of presentation order on user judgments of document relevance. *Journal of the American Society for Information Science* 39(5):293 – 300.
- Ghahramani, Z. 1998. Learning dynamic bayesian networks. In *Adaptive Processing of Sequences and Data Structures*, 168–197. Springer-Verlag.
- Hou, Y., and Song, D. 2009. Characterizing pure high-order entanglements in lexical semantic spaces via information geometry. In *Proceedings of the Third Quantum Interaction Symposium*, 237–250.
- Khrennikov, A. 2004. On quantum-like probabilistic structure of mental information. *Open Systems and Information Dynamics* 11(3):267–275.
- Kittur, A.; Chi, E. H.; and Suh, B. 2008. Crowdsourcing for usability: Using micro-task markets for rapid, remote, and low-cost user measurements. In *ACM CHI 2008*, 453–456.
- Kondacs, A., and Watrous, J. 1997. On the power of quantum finite state automata. In *Proceedings of the 38th IEEE Conference on Foundations of Computer Science*, 66–75.
- Melucci, M. 2008. A basis for information retrieval in context. *ACM Trans. Inf. Syst.* 26(3):1–41.
- Melucci, M. 2010. An investigation of quantum interference in information retrieval. In *IRF 10*, In press.
- Mizzaro, S. 1996. Relevance: The whole (hi)story. *Journal of the American Society for Information Science* 48:810–832.
- Piwovarski, B., and Lalmas, M. 2009a. Structured information retrieval and quantum theory. In *Proceedings of the Third Quantum Interaction Symposium*, 289–298.

- Piwowarski, B., and Lalmas, M. 2009b. A quantum-based model for interactive information retrieval. In *Proceedings of the 2nd International Conference on the Theory of Information Retrieval*, volume 5766. Springer.
- Tzeng, W.-G. 1992. Learning probabilistic automata and markov chains via queries. *Machine Learning* 8:151–166.
- van Rijsbergen, C. J. 2004. *The Geometry of Information Retrieval*. New York, NY, USA: Cambridge University Press.
- Vidal, E.; Thollard, F.; de la Higuera, C.; Casacuberta, F.; and Carrasco, R. C. 2005. Probabilistic finite-state machines-part i. *IEEE Trans. Pattern Anal. Mach. Intell.* 27(7):1013–1025.
- Zuccon, G., and Azzopardi, L. 2010. Using the quantum probability ranking principle to rank interdependent documents. In *Proceedings of the 32nd European Conference on Information Retrieval*, 357–369.
- Zuccon, G.; Azzopardi, L.; and van Rijsbergen, C. J. 2009. The quantum probability ranking principle for information retrieval. In *Proceedings of the 2nd International Conference on the Theory of Information Retrieval*, 232–240.