Explanation of Relevance Judgement Discrepancy with Quantum Interference

Jun Wang[1], Dawei Song[1], Peng Zhang[1], Yuexian Hou[2], Peter Bruza[3]
[1]School of Computing, The Robert Gordon University, UK.
[2]School of Computer Science & Technology, Tianjin University, China.
[3]Faculty of Science and Technology, Queensland University of Technology, Australia
Email: j.wang3, d.song, p.zhang1@rgu.ac.uk,yxhou@tju.edu.cn,p.bruza@qut.edu.au

Abstract

A key concept in many Information Retrieval (IR) tasks, e.g. document indexing, query language modelling, aspect and diversity retrieval, is the relevance measurement of topics, i.e. to what extent an information object (e.g. a document or a query) is about the topics. This paper investigates the interference of relevance measurement of a topic caused by another topic. For example, consider that two user groups are required to judge whether a topic q is relevant to a document d, and q is presented together with another topic (referred to as a companion topic). If different companion topics are used for different groups, interestingly different relevance probabilities of q given d can be reached. In this paper, we present empirical results showing that the relevance of a topic to a document is greatly affected by the companion topic's relevance to the same document, and the extent of the impact differs with respect to different companion topics. We further analyse the phenomenon from classical and quantum-like interference perspectives, and connect the phenomenon to nonreality and contextuality in quantum mechanics. We demonstrate that quantum like model fits in the empirical data, could be potentially used for predicting the relevance when interference exists.

Introduction

The measurement of topic relevance is important in many IR tasks. During the relevance judgement process, we spotted an interesting phenomenon. That is, two user groups often can not reach an agreement on the same topic when they also judge a different additional topic at the same time.

For instance, consider two topics "Brave Heart" (William Wallace's nickname and the name for his film biography) and "William Wallace", and a biographical article about William Wallace. Both topics are topically relevant to the article. However, "Brave Heart" may possibly be judged as irrelevant when it is displayed together with a more relevant topic "William Wallace" (as a companion topic), while if "Brave Heart" is shown together with a completely irrelevant topic, it might be judged with high relevance probability.

We speculate that relevance of one topic to a document is affected by the companion topic's relevance judged on the same document. If the hypothesis is true, it can have important theoretical and practical implications. For example, when the documents are being indexed by a list of topics (in the form of, e.g. keywords or concepts), the topic weights should be adjusted according to surrounding topics in the documents in order to better represent their relevance. Take query expansion as another example, which reformulates the query with synonyms or terms statistically related to a set of relevant documents. When the suggested query terms are presented together for users to select, due to the possible topic interference caused by the companion terms, their actual relevance degrees to the original query may largely deviate from the scores assigned to them.

In this paper we will empirically and theoretically investigate such kind of relevance judgement discrepancy and try to quantify the interference from classical and quantum-like points of view. There are a few reasons why we are seeking quantum like modelling for explaining such relevance discrepancy.

Firstly, the relevance of a topic is actualized during the retrieval time, which is similar to the inherent non-reality property of quantum mechanics: A superposed topic with a few possible meanings could yield any of them. Only when the users interact with the topic and a document within certain context, its meaning will collapse to a specific state, then the relevance of the topic will come into reality. However, such contextualization only happens when the relevance judgement, i.e. measurement, happens. This coincides with the non-reality principle of quantum mechanism.

Also the recent work in modelling the probability judgement error in psychology (Busemeyer, Franco, and Pothos 2009) and words association in cognitive science (Bruza and Woods 2008) inspired us to look into the quantum-like explanation of the relevance discrepancy problem in IR, as relevance judgements in IR can also be viewed as a kind of social-psychological behaviour. For example when a user is looking at queries or documents, he or she will unconsciously associate certain words with the others, and the user's cognitive state evolves with respect to what he or she has read. This in turn affects the user' relevance judgements.

In the next section, we give a brief introduction to the related work in quantum-like phenomena outside physics.

Related Work

Quantum mechanics was originally used to describe particles' state and evolution in micro-world, and has been extensively studied and applied in many other areas such as Biology, Sociology and Psychology etc. (Busemeyer, Franco, and Pothos 2009; Baaquie 2004; Patel 2001; Hameroff 1994).

In Cognitive Science, (Bruza and Cole 2005; Bruza and Woods 2008) studied the free association of words and showed that the formalism of superposed state in quantum mechanics can be used to model the fact that words having multiple meanings tend to collapse to one meaning according to the context in semantic spaces. The entanglements assumption in modelling free word associations can explain the different strength of association among the terms. (Bruza et al. 2009).

In Psychology, the judgement probability fallacy has long been discovered, and classical probability theory fails to properly explain the fallacy (Tversky and Shafir 1992). Recently, researchers successfully discovered the cause of the fallacies with quantum-like representations (Franco 2007; Busemeyer and Wang 2007; Aerts 2007; Aerts and Gabora 2005; Aerts, Aerts, and Gabora 2009). It was found that the violation of the classical probability theory is rooted on that fact that the subjects are uncertain about the concept. The more uncertain the subjects are, the more likely the judgement probability will violate the classical logic.

Meanwhile, quantum mechanics has constantly attracted interests from IR researchers. van Rijsbergen advocated the application of quantum theory to IR (van Rijsbergen 2004), and pointed out the suitability of quantum mechanics concepts and formalisms (e.g. Hilbert space, superposition, Hermitian operator) in representing and measuring the uncertainty in IR. It is believed that quantum theory can bring together current existing IR ranking models into one single framework, enable formal analysis, and embed searching context as well as user information needs into the retrieval model.

In the past few years, some quantum mechanics based theoretical IR models have emerged. For example, Melucci explored a method for modeling context and ranking object with context (Melucci 2008; 2005). Piwowarski and Lalmas (Piwowarski and Lalmas 2009) proposed to model user's information need as a state vector in Hilbert space. Hou and Song (Hou and Song 2009) proposed an expanded vector space model that involves non-separable (pure) high order word associations. Zuccon et al. (Zuccon, Azzopardi, and van Rijsbergen 2009) proposed to incorporate the interference between documents in the ranking scheme, aiming to circumvent the independence assumption in the traditional probabilistic ranking principle. Nevertheless, less attention has been paid to research on the quantum-like interference between topics in the relevance judgement process, which is increasingly important to many IR tasks such as document indexing, query language modelling, aspect and diversity retrieval.

In the next section, we will present some initial empirical evidence for the topic interference phenomenon, and show that the phenomenon can not be fully explained from a classical perspective.

A Statistical Evidence of Relevance Judgement Discrepancy

Experimental Settings

The aim of this user experiment is to test our topic interference hypothesis: the relevance of one topic can enhance or reduce the probability of judging another topic to be relevant to a document.

In this experiment, we designed 10 groups of topics in total. Each group contains three topics A, B and C, which have some sort of connections to each other. For example, in the first group, the three topics are "Brave Heart", "William Wallace" and "William's wife", where "Brave Heart" is the nickname of "William Wallace", and "William's wife" refers to the wife of "William Wallace". As another example, in topic group 3, the three topics "probability", "distribution" and "statistics" are closely related concepts. When one of them is mentioned, people always recall the other two if they have a reasonable level of knowledge about probability and statistics. We list all the 10 topic groups in Table 1.

For each topic group, we picked two documents from Google searching results. The query for Google search is the conjunction of topics A, B and C. Thus each document is more or less about topic A, B or C. The two selected documents are different in detailed content, but at least some terms from topic A, B and C appear in the documents.

We allowed each user to see one and only one randomly assigned pair of topics (either A and B, or A and C, or B and C) simultaneously when he or she was asked to do the relevance judgement. Each user was only allowed to participate once in the experiment in order to rule out the user cognitive state change (e.g., background knowledge due to learning effect) due to the previous judgements. And they were required to do all the ten topic groups without knowing our topic interference hypothesis.

Some users may misunderstand that there is only one topic relevant to a document when they can see two at the same time and then they would purposely choose one topic as irrelevant. Such misunderstanding could make the experiment end up with more topics judged as irrelevant and intensify the topic interference. To eliminate such possibility, we stated explicitly in experiment guideline: *For each document, there can be one, or two, or none of the topics which are relevant to the document..*

The experiment was carried out on Mechanical Turk[1], and the online interface is shown in Figure 1. Mechanical Turk is an Amazon Web Service, which enables online user studies to be "crowd-sourced" to a large number of users quickly and economically. When the experiment are properly designed, the quality of data collected from Mechanical Turk can be as good as those from evaluation experts (Alonso and Mizzaro 2009).

To gain a better controllability of user experiments on Mechanical Turk, it is important to detect suspicious responses

[1] https://www.mturk.com/mturk/welcome

Topic group	A	B	C
1	Brave heart	William Wallace	William's wife
2	Beijing Olympic 2008	Bolt 100m world record	Athlete, Bolt training facility
3	Probability	Distribution	Statistics
4	US army	Barack Obama	Laura, US first lady, daily life
5	ACM Turing award	Computer Science Professor	Prof. Qiang Yang at HKUST
6	Oil plant energy	Aberdeen	North sea
7	Semantic web	Ontology	Theology
8	Young girls luxury	Fashion	Outdoor sports
9	Relocation company	Property leasing	Packing box
10	Bach	Power law	1/f noise

Table 1: All the topics in our topic interference experiment

Figure 1: Interface of the interference experiment

(Kittur, Chi, and Suh 2008). Here we recorded users' click information to detect the "unreliable" users, e.g., those who did not complete all the 10 topic groups, or simply judged all the documents as relevant or irrelevant in a very short time, or have not clicked the links leading to the documents. These users either have not fulfilled experiment, or did not read the document. Therefore, we can not rely on the results of their work. Finally we have gathered the results from 80 users in total. After removing the results from the detected "unreliable" users, we obtained 59 valid results: 18 from the group with topic A and B, 22 from the group with topic B and C, and 19 from group with topic A and C. The probability of a topic being relevant to a document is computed as the ratio of the number of users who judged the topic as relevant to the total number of users who judged on it.

Experimental Results and a Classical Explanation

We list the relevance probabilities of all the topics to their corresponding documents and their companion topics in Table 2. It shows there is a more or less disagreement (Δ) on the relevance of each topic to a document, when being displayed together with a different companion topic. We plot in Figure 2 the variance of relevance probabilities for all topics when they are displayed with the other two topic separately. The X-axis stands for the different topic groups, and Y-axis stands for the relevance probability. For instance, each vertical line in the Figure 1(A) is the gap of two relevance probabilities for topic A when displayed with topic B and C respectively. The Y coordinate for each end of the vertical line are the real probabilities for topic A being relevant to the document when displayed with B and C respectively. Each pair of closely aligned vertical lines represent the judgements on the same topic group, but on different documents.

Let us consider the earlier example "Brave Heart" (topic A in group 1), which has a distinctive relevance discrepancy on the document (6449), an "William Wallace" biography article from Wikipedia. This article contains some information about "William Wallace" biography and Scotland history. The film "Brave heart" based on the story of William Wallace is mentioned at the end of the article. However, William's wife does not appear in any place of the article.

The details of user relevance judgements for each pair of topics from this group are given in Table 3. The rel-

Topic Group	DocID	P(A)			P(B)			P(C)		
		B	C	Δ	A	C	Δ	A	B	Δ
1	6448	0.71	0.79	-0.08	0.79	1.00	-0.21	0.11	0.11	0.00
1	6449	0.14	0.89	-0.75	0.93	1.00	-0.07	0.05	0.11	-0.06
2	6450	0.79	0.74	0.05	0.71	0.89	-0.18	0.53	0.00	0.53
2	6451	0.07	0.42	-0.35	0.78	0.67	0.11	0.42	0.11	0.31
3	6461	0.93	0.95	-0.02	0.64	1.00	-0.36	0.63	0.89	-0.26
3	6462	0.93	0.90	0.03	0.43	0.89	-0.46	0.64	1.00	-0.36
4	6452	0.50	1.00	-0.50	0.93	1.00	-0.07	0.00	0.00	0.00
4	6453	0.36	0.42	-0.06	0.57	1.00	-0.43	0.21	0.00	0.21
5	6454	1.00	1.00	0.00	0.07	0.33	-0.26	0.00	0.00	0.00
5	6455	0.57	0.79	-0.22	1.00	0.78	0.22	0.11	0.11	0.00
6	6463	0.42	0.68	-0.26	0.64	1.00	-0.36	0.05	0.00	0.05
6	6464	0.28	0.27	0.01	0.71	0.78	-0.07	0.69	0.89	-0.20
7	6457	0.21	0.21	-0.00	0.79	1.00	-0.21	0.42	0.33	0.09
7	6458	0.93	0.89	0.04	0.14	0.44	-0.30	0.05	0.11	-0.06
8	6459	0.14	0.16	-0.02	0.79	0.78	0.01	0.89	0.89	0.00
8	6460	0.57	0.74	-0.17	0.86	0.89	-0.03	0.11	0.11	0.00
9	6465	0.78	0.84	-0.06	0.28	0.56	-0.28	0.16	0.11	0.05
9	6467	1.00	0.89	0.11	0.14	0.22	-0.08	0.32	0.78	-0.46
10	6456	0.07	0.00	0.07	0.57	0.44	0.13	0.53	0.44	0.09
10	6466	0.36	0.31	0.05	0.58	0.33	0.25	0.10	0.11	-0.01

Table 2: Variance of relevance probability of topics when displayed with another topic

evance probabilities of topic "Brave Heart" (A) are 0.14 ($P(A,B)$+$P(A,\bar{B})$) and 0.89 ($P(A,C)$+$P(A,\bar{C})$) respectively, when it is displayed with "William Wallace" (B) and "William's wife" (C) separately. It is interesting that the variance "Brave Heart"'s relevance probability on the same document can be so large. A closer analysis suggests that disagreements in topic relevance are more likely to stem from the comparison effect between topics, rather than from the uncertainty associated with information in the underlying document. This comparison effect is not the same as what has been discussed in (Huang and Soergel 2006) where users compared the evidence from document in order to decide whether a document is relevant to the topic. Here it is more about the degree of a topic relevant to a document.

When the topics "Brave Heart" (A) and "William Wallace" (B) were displayed together, 93% of the users chose to judge the document as relevant to "William Wallace", 14% of the users judged the document as relevant to "Brave Heart". One possible explanation for the discrepancy in relevance probability between "Brave Heart" and "William Wallace" is that "Brave Heart" is mentioned at the end of the document and perhaps many users did not see this. However this can be ruled out because when "Brave Heart" was displayed together with "William's wife" (C), 89% of the users judged it as relevant, while only 5% of the users judged "William's wife" as relevant.

Why did the relevance probability for "Brave Heart" largely increase, in comparison with the 14% when it was displayed with "William Wallace"? We think it is most likely that people tended to choose "Brave Heart" as relevant when they could not find out the information about "William's wife", although they normally would not think "Brave Heart" is very relevant (as verified in the first case

A. Brave Heart; B. William Wallace; C. William's wife

William Wallace (http://en.wikipedia.org/wiki/William_Wallace)

AB	A	$\bar{A}$
B	0.07	0.86
$\bar{B}$	0.07	0.00

BC	B	$\bar{B}$
C	0.11	0.00
$\bar{C}$	0.89	0.00

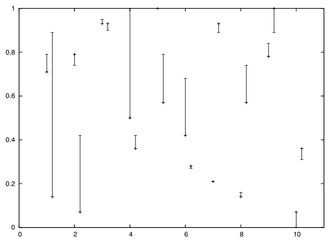
AC	A	$\bar{A}$
C	0.05	0.00
$\bar{C}$	0.84	0.11

Table 3: Relevance judgement on document "William Wallace" from each topic pair

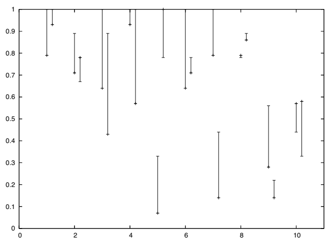
when it was displayed together with "William Wallace") and they have been informed that there can be the case that none of the topics is relevant to the document. Therefore, we believe that the comparison effect against the low baseline ("William's wife") may be the most possible reason for this disagreement. Similarly, the high relevance probability of "William Wallace" interferes the relevance judgement for "Brave Heart" by setting a high comparison baseline, which results in the low relevance probability of "Brave Heart".

The same phenomena can be observed for other topic groups. We found that the disagreement is also related to the specificity of a document regarding that topic. The less specific the document to the topic, the higher disagreement in the relevance judgement. At this stage, we also do not rule out the possibility that some users may lack the knowledge about the topics, leading to some degree of uncertainty in making topical relevance decisions.

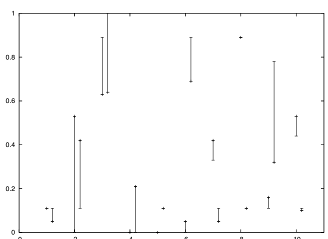
However, we do agree that the interference may not exist in some circumstances. For example if two topics are



(a) P(A)



(b) P(B)



(c) P(C)

Figure 2: Disagreement of relevance probabilities of a topic when displayed another topic

largely overlapped, as “probability” and “distribution” with “statistics”, then the interference could be barely observed. Or if the two topics are totally irrelevant, i.e. one of which is about the document with great certainty and the other one is completely has nothing to do with the document, then little interference could be observed either. One example for this case is “semantic web” and “ontology” with “theology”,

The above intuitive comparison explanation fits into the empirical results, yet it is difficult to give the reason why some interferences are high and some are low. In the next section we will use a quantum like model to quantitatively explain our relevance judgement result.

Explanation with Quantum-like Modelling

Basic Quantum Representation

Let us start with the binary relevant judgement, in which there are only two outcomes: relevant and irrelevant. Then a document can be represented by state $|q\rangle$ and $|\bar{q}\rangle$, meaning the document is about a query or not about a query respectively. The superposed state of a relevance judgement between a topic q and a document d is:

$$|d\rangle = a_1|q\rangle + a_2|\bar{q}\rangle \tag{1}$$

where $a_1^2 + a_2^2 = 1$, and a_1^2 is the probability of topic q being relevant and a_2^2 is the probability of topic q being irrelevant. When a_j $(j = 1, 2)$ is a complex number $\alpha_j cos\theta + i\alpha_j sin\theta$, it can also be written as $a_j e^{i\theta}$ which has the property $a_j^2 = (\alpha_j e^{i\theta_j})^* \alpha_j e^{i\theta_j} = \alpha_j^2$.

Quantum Contextualization

In cognitive science, Barsalou (Barsalou 1982) found that a concept contains two types of properties: context-independent properties activated by the words in all occasions, and context-dependent properties activated by the related context in which it appears. Potential concept of an system changes in different ways when the system interacts with different contexts. The context may also change with respect to its interaction with the system. Then the system undergoes another interaction, and so forth.

Similarly in the relevance judgement in IR, whether a document is relevant to a topic depends largely on the user’s knowledge state, user’s information need, as well as the environment where the judgement is carried out. All of them are considered as retrieval context, and anything users perceived during the judgement process will also be ascribed as a context. Before users interact with the retrieval system, the document is either relevant to the topic or not. When the user starts the judgement, the relevance comes to play. This process can be viewed as concept contextualization.

Consider the “Brave Heart” example again. It has two different relevance probabilities when it is displayed with “William Wallace” and “William’s wife” respectively. Here “William Wallace” and “William’s wife” can be viewed as two different contexts, the relevance judgement on “Brave Heart” can be viewed as contextualization.

Following Khrennikov’s hyperbolic quantum formalism and context dependent probability (Khrennikov 2010; 2009), we assume the companion topic q_1 to be a context, then the document can be described by the basis of context c_{q_1}

$$|d\rangle = b_1|q_1\rangle + b_2|\bar{q_1}\rangle \tag{2}$$

where $b_1^2 + b_2^2 = 1$, $b_j = \beta_j e^{i\theta_j}$, and $|q_1\rangle$ and $|\bar{q_1}\rangle$ is the orthogonal basis of context c_{q_1}.

If query $|q\rangle$ is also represented with the basis $|q_1\rangle$ and $|\bar{q_1}\rangle$, then:

$$|q\rangle = b_3|q_1\rangle + b_4|\bar{q_1}\rangle \tag{3}$$

where $b_3^2 + b_4^2 = 1$, $b_k = \beta_k e^{i\theta_k}$. The probability interpretation for the numbers are: $b_3^2 = P(q|q_1)$, $b_4^2 = P(q|\bar{q_1})$.

The inner product between q and d gives the probability that a document in state d will be found in state q, which can be viewed as their relevance probability:

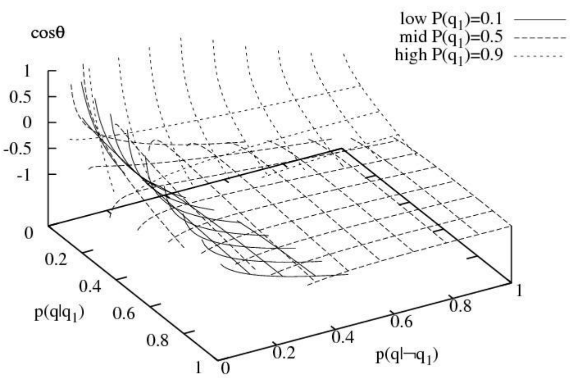


Figure 3: $P(q) = 0.1$, low

$$
\begin{aligned}
P(q|d) &= |<q|d>|^2 & (4)\\
&= (b_1b_3 + b_2b_4)^*(b_1b_3 + b_2b_4) & (5)\\
&= \beta_1^2\beta_3^2 + \beta_2^2\beta_4^2 & (6)\\
&+ 2\beta_1\beta_3\beta_2\beta_4 cos((\theta_2 - \theta_1) + (\theta_4 - \theta_3))
\end{aligned}
$$

we have the following probabilities.

$$
\begin{aligned}
P_{q_1}(q|d) &= P(q|q_1)P(q_1|d) + P(q|\bar{q_1})P(\bar{q_1}|d) & (7)\\
&+ 2cos\theta(q, q_1, d)\sqrt{\prod_{q_1=r,\bar{r}} P(q|q_1)P(q_1|d)}
\end{aligned}
$$

where $P(q|q_1)$ is the probability that topic q is relevant within context c_{q_1}, $P(\bar{q}|q_1)$ is the probability that topic q is irrelevant within context c_{q_1}, $P(q_1)$ is the probability that context c_{q_1} occurs, and $P(\bar{q_1})$ is the probability that context c_{q_1} does not occur. $\theta(q, q_1, c) = (\theta_1 - \theta_2) + (\theta_3 - \theta_4)$.

From Equation 7, the interference happens between probabilities from different contexts, and also the probability of context. We can also see that when $P(q_1)$ or $P(\bar{q_1})$ is 0, which means there is no uncertainty with the context c_{q_1}, the decision on topic "q" is interfered by the uncertainty involved in the context. This explanation is different from classical total probability where the relevance probability of q purely comes from the two probabilities from two possible context outcomes: $P(q) = P(q|q_1)P(q_1) + P(q|\bar{q_1})P(\bar{q_1})$.

Even if the relevance of q is independent of q_1, there still exists a possibility of interference. If q is displayed with another topic q_2, which has the same probabilities as q_1, it still can not be guaranteed that the relevance of q will be judged as the same as before,due to the difference of phase angles of companion topics.

Next let us visualize the interference from the phase angles. In the quantum like model, there are many factors which influence the interference parameter $cos\theta$. It is difficult for us to plot the tendency regarding all the priors. Therefore, we choose the fixed $P(q_i)$ with high (0.9), middle (0.5) and low (0.1) values respectively, and do so for the

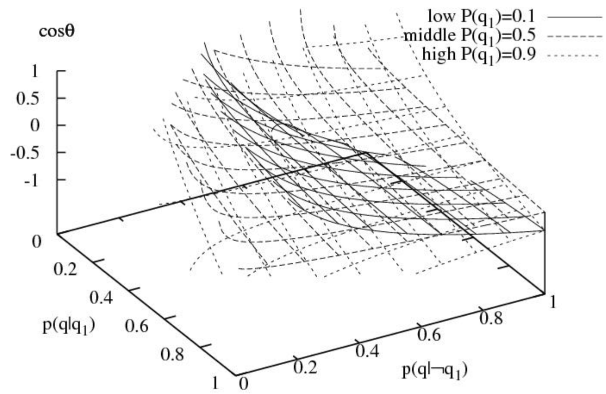


Figure 4: $P(q) = 0.5$, middle

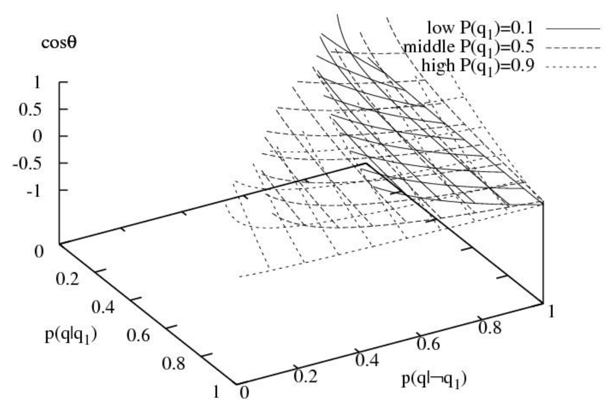


Figure 5: $P(q) = 0.9$, high

decision on topic q. But the conditional probability continuously changes from 0 to 1. From Figures 3,4,5, we can make following observations: When $P(q_1)$ and $P(q|q_1)$ are both small, the interference is high. When $P(q_1)$ and $P(q|q_1)$ are high, the interference from two topics is negative, leading to decrease of $P(q)$. These tendency graphs actually have explained the phenomenon that "Brave Heart" has varied relevance probabilities when displayed with highly or lowly relevant companion topics.

According to the definition of $cos\theta$, we can draw the same conclusion. Suppose we have two topics q_1 and q_2 displayed with q separately, the relevance probability of q_1 is smaller than q_2 (i.e. $P(q_1) < P(q_2)$), and both of them are not higher than 0.5. Also suppose the conditional probabilities are the same (i.e. $P(q|q_1) = P(q|q_2)$). Then the ratio of two $cos\theta$ based on the two companion topics q_1 and q_2 will be:

$$\frac{cos\theta_{q_1}}{cos\theta_{q_2}} = \frac{P_{q1}(q) - P(q, q_1) - P(q, \bar{q_1})}{P_{q2}(q) - P(q, q_2) - P(q, \bar{q_2})} \cdot \frac{\sqrt{\prod_{q_2=r,\bar{r}} P(q, q_2)}}{\sqrt{\prod q_1 = r, \bar{r} P(q, q_1)}} \tag{8}$$

Because $P(q_1) < P(q_2) < 0.5$, the second part of Equation 8 is greater than 1. If the conditional probability $P(q|q_j)$ is greater than $P(q|\bar{q_j})$, i.e. the judgement of q is more relying on the relevance, instead of irrelevance, of q_j, then the first part of equation 8 is also greater than 1 (as $P(q_2) \cdot P(q_|q_2) + (1 - P(q_2)) \cdot P(q|\bar{q_2}) > P(q_1) \cdot P(q|q_1) + (1 - P(q_1)) \cdot P(q|\bar{q_1})$). As a result, the interference in context q_1 is stronger than q_2. To simplify the analysis, we have to assume some probabilities to be constant, but the above formula explains to certain extent why some topics are judged to be of higher relevance when they are shown with a lowly relevant companion topic. The reason why some topics are judged as less relevant when they are shown with a more relevant companion topic can be explained in the same way.

Conclusion

In this paper, we have shown initial empirical evidence that relevance is not an intrinsic feature of a document. Only when users interact with a topic and a document, the relevance is realized with respecting to the context. When there is companion topic, the relevance of the companion topic and the distance between the two topics both influence the topic relevance judgement.

The relevance judgement is a rather complicated process. It is difficult to analyse exactly what factors, when and to which extent are playing a role here. As we can see that changing one component in the relevance judgement environment, for instance, the companion topic, will end up with largely different result. The explanation via Quantum-like interference shows that the distance of two topic and the relevance of the companion topic can significantly influence the topic relevance. Successful modelling such interference in retrieval models will be likely to have a fundamental impact on the IR field. For example, any task involving topic relevance measurement would benefit from the models incorporating the interference of the topics and their context.

Acknowledgements

We are grateful to reviewers for their constructive comments. This work is funded in part by the UK's Engineering and Physical Sciences Research Council (grant number: EP/F014708/2).

References

Aerts, D.; Aerts, S.; and Gabora, L. 2009. Experimental evidence for quantum structure in cognition. In *QI '09*, 59–70. Berlin, Heidelberg: Springer-Verlag.

Aerts, D., and Gabora, L. 2005. A theory of concepts and their combinations: The structure of the sets of contexts and properties. *Kybernetes* 34:167–191.

Aerts, D. 2007. Quantum interference and superposition in cognition: Development of a theory for the disjunction of concepts. arXiv:0705.0975v2.

Alonso, O., and Mizzaro, S. 2009. Can we get rid of trec assessors? using mechanical turk for relevance assessment. SIGIR IR Evaluation Workshop.

Baaquie, B. E. 2004. *Quantum Finance: Path Integrals and Hamiltonians for Options and Interest Rates*. Cambridge University Press.

Barsalou, L. W. 1982. Context-independent and context-dependent information in concepts. *Memory & Congnition* 10(1):82–93.

Bruza, P., and Cole, R. J. 2005. Quantum logic of semantic space: An exploratory investigation of context effects in practical reasoning. In *We Will Show Them! (1)*, 339–362.

Bruza, P., and Woods, J. 2008. Quantum collapse in semantic space : interpreting natural language argumentation. In *QI'08*.

Bruza, P.; Kitto, K.; Nelson, D.; and McEvoy, C. 2009. Extracting spooky-activation-at-a-distance from considerations of entanglement. In *QI'09*, 71–83.

Busemeyer, J. R., and Wang, Z. 2007. Quantum information processing explanation for interactions between inferences and decisions. In *QI'07*.

Busemeyer, J. R.; Franco, R.; and Pothos, E. M. 2009. Quantum probability explanations for probability judgment 'errors'. eprint arXiv:0909.2789.

Franco, R. 2007. The conjunction fallacy and interference effects. *Journal of Mathematical Psychology*.

Hameroff, S. R. 1994. Quantum coherence in microtubules: A neural basis for emergent consciousness? *Journal of Consciousness Studies* 1(1):91–118.

Hou, Y., and Song, D. 2009. Characterizing pure high-order entanglements in lexical semantic spaces via information geometry. In *QI'09*, 237–250.

Huang, X., and Soergel, D. 2006. An evidence perspective on topical relevance types and its implications for exploratory and task-based retrieval. In *ISIC*.

Khrennikov, A. Y. 2009. *Contextual Approach to Quantum Formalism*. Fundamental Theories of Physics. Springer.

Khrennikov, A. Y. 2010. *Ubiquitous Quantum Structure: From Psychology to Finance*. Springer.

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 *Proceedings of the 26th Annual ACM Conference on Human Factors in Computing Systems (CHI2008)*, 453–456.

Melucci, M. 2005. Context modeling and discovery using vector space bases. In *CIKM '05*, 808–815. New York, NY, USA: ACM.

Melucci, M. 2008. A basis for information retrieval in context. *ACM Transactions on Information Systems* 26(3):1–41.

Patel, A. 2001. Why genetic information processing could have a quantum basis. *Journal of Biosciences*.

Piwowarski, B., and Lalmas, M. 2009. Structured information retrieval and quantum theory. In *QI '09*, 289–298. Berlin, Heidelberg: Springer-Verlag.

Tversky, A., and Shafir, E. 1992. The disjunction effect in choice under uncertainty. *Psychological Science*.

van Rijsbergen, C. J., ed. 2004. *The Geometry of Information Retrieval*. Cambridge University Press.

Zuccon, G.; Azzopardi, L.; and van Rijsbergen, K. 2009. The quantum probability ranking principle for information retrieval. In *ICTIR'09*, 232–240.