The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Exploring Relevance Judgement Inspired by Quantum Weak Measurement

Tianshu Wang, Yuexian Hou,* Panpan Wang, Xiaolei Niu
Tianjin University, Tianjin, China

Abstract

Quantum Theory (QT) has been applied in a number of fields outside physics, e.g. Information Retrieval (IR). A series of pioneering works have verified the necessity to employ QT in IR user models. In this paper, we explore the process of relevance judgement from a novel perspective of the two state vector quantum weak measurement (WM) by considering context information in time domain. Experiments are carried out to verify our arguments.

Introduction

The definition of relevance is a key concept in IR. Most of the traditional works in IR are based on the static relevance, which is to judge whether a document is relevant to a query. However, real retrieval process is more complex. Human-based relevance is important to employ effective user models.

(Wang et al. 2016) first explored the quantum-like phenomenon when judging relevance in IR from a real user study perspective. By user experiments, they found the violation of the law of total probability and the validity of the order effect in real user's judgement. They proposed that the relevance judgement triggers the change of user's IN in IR.

However, there are some special examples in IR which can not be explained well using the standard QT. The quantum WM (Aharonov and Vaidman 2008) is the generalization of the standard quantum measurement and we consider that the WM is more suitable to model the mysterious cognitive IR process. In this paper, to verify our assumption, both user experiment and experiments on the session search task are carried out. The results verify our arguments.

Background

Imagine in this scenario that there are two groups of users to judge two documents d_A and d_B (Wang et al. 2016). Participants from the first group should judge d_A and then d_B, while participants from the second group should judge d_B only. We denote r_A as the event that a participant judges d_A as "relevant", $\overline{r_A}$ as the event that a participant judges d_A as "not relevant" and similarly for r_B and $\overline{r_B}$. $p'(r_B)$ means that the probability of the first group who judge d_B as "relevant" after judging d_A and $p(r_B)$ means that the probability of the second group who directly judge d_B as "relevant". $\{r_A, \overline{r_A}\}$ and $\{r_B, \overline{r_B}\}$ form two measuring bases.

In classical probability, because of the law of total probability, $p'(r_B) = p(r_A) * p(r_B|r_A) + p(\overline{r_A}) * p(r_B|\overline{r_A})$. Furthur, $p(r_B)$ should equal to $p'(r_B)$.

In standard quantum probability, events are defined in a Hilbert space and represented by a column vector $|\phi\rangle$. The conjugate transpose of $|\phi\rangle$ is $\langle\phi|$. The users' cognitive state may be demoted as a superposed state:

$$|S\rangle = \alpha|r_A\rangle + \beta|\overline{r_A}\rangle \tag{1}$$

where $|\alpha|^2 + |\beta|^2 = 1$. α and β are the amplitudes. If we use a projector $|r_A\rangle\langle r_A|$ to measure $|S\rangle$, we have the probability of $|\alpha|^2$ to have $|r_A\rangle$ and vice versa. The concept of superposition can represent user's inner conflict during judging vividly and after the judgement, $|S\rangle$ would collapse to one of the basis states ($|r_A\rangle$ or $|\overline{r_A}\rangle$).

$$p'(r_B) = |\langle q|r_A\rangle|^2 * |\langle r_A|r_B\rangle|^2 + |\langle q|\overline{r_A}\rangle|^2 * |\langle\overline{r_A}|r_B\rangle|^2 \tag{2}$$

where $|q\rangle$ is the representation of query and $p(r_B)$ is:

$$\begin{aligned} p(r_B) =& |\langle q|r_A\rangle|^2 * |\langle r_A|r_B\rangle|^2 + |\langle q|\overline{r_A}\rangle|^2 * |\langle\overline{r_A}|r_B\rangle|^2 \\ &+ 2|\langle q|r_A\rangle\langle r_A|r_B\rangle\langle q|\overline{r_A}\rangle\langle\overline{r_A}|r_B\rangle|cos(\theta) \end{aligned} \tag{3}$$

where θ is the angle between two measuring bases(Wang et al. 2016). So in quantum probability, $p(r_B)$ doesn't always equal to $p'(r_B)$.

(Wang et al. 2016) carried out a real user experiment. The result shew that $p(r_B)$ didn't always equal to $p'(r_B)$, which agreed with the quantum probability. It also proved the existence of order effect in relevance judgement.

Quantum Weak Measurement

Assume that we have a system of Equation 1. In physics, the quantum measurement is performed by interacting the measured system with the measuring device. In general, the probability distribution of the position of measuring device pointer is modeled by Gaussian function (Aharonov and Vaidman 2008). After the interaction, the probability distribution of the pointer variable is:

$$\begin{aligned} P(Q) =& (2\pi\Delta^2)^{-1/2}(|\alpha|^2 e^{-(x-V_{r_A})^2/2\Delta^2} \\ &+ |\beta|^2 e^{-(x-V_{\overline{r_A}})^2/2\Delta^2}) \end{aligned} \tag{4}$$

*Corresponding author: Yuexian Hou (yxhou@tju.edu.cn)

where x is the position of the pointer and Δ^2 is the variance. V_{r_A} and $V_{\overline{r_A}}$ are the corresponding eigenvalues of r_A and $\overline{r_A}$ in $|S\rangle\langle S|$. Generally speaking, Δ^2 in WM is much larger than in standard measurement.

In standard measurement, because the Δ^2 is small, the pointer would fall on V_{r_A} or $V_{\overline{r_A}}$ with high probability. If the pointer falls on V_{r_A}, $|\beta|^2 e^{-(x-V_{\overline{r_A}})^2/2\Delta^2}$ will approximately equal to zero while $|\alpha|^2 e^{-(x-V_{r_A})^2/2\Delta^2}$ would keep effective (and vice versa). In this way, the system will collapse to the corresponding basis state.

Then, if the Δ^2 becomes much larger, the pointer may not fall on a range of values (not just close to V_{r_A} or $V_{\overline{r_A}}$). What's more, both $|\alpha|^2 e^{-(x-V_{r_A})^2/2\Delta^2}$ and $|\beta|^2 e^{-(x-V_{\overline{r_A}})^2/2\Delta^2}$ may all keep effective after the measurement. In this situation, the system will not collapse to the basis state but bias a little, which means that the system is still superposed.

(Aharonov and Vaidman 2008) gave a formula to calculate the statistical result of weak measurement in the two-state-vector formalism, called weak value:

$$W = \frac{\langle\phi|\, A\, |\psi\rangle}{\langle\phi|\psi\rangle} \tag{5}$$

where $|\psi\rangle$ and $\langle\phi|$ is the past and future state of the system respectively. A is an observable (in this paper $A = |S\rangle\langle S|$). The weak value is defined by considering context information in time domain and modeling various coupling strength between systems and measurement devices.

User Experiment

We recruited 20 participants and carried out a user experiment. In the experiment, each person was asked to judge the relevance of a query-document pair with a score from -4 to 4. The higher the score is, the more relevant the document is. What we concerned is whether they would change their answers after the first judgement.

The result is that: (1) there is an apparent judging discrepancy across different users; (2) the larger the variance of judgement is, the more likely the users change their first judgement. Relevance judgement is a cognitive activity, so it's hard to be measured precisely. The meaning of the scores are vague and discrepant (like the big variance in WM), which may lead to the discrepancy of final judgement. From the perspective of standard measurement, after the user gives his answers, his cognitive state would collapse to the corresponding state of the answer and the conflict (or superposition) would vanish. Then no matter how more times asked to make the same judgement, the user's cognition will not change and the answer would keep consistent, which is against the result in our experiment. Even after the users make the judgement, their cognition may be still superposed (like WM). These cause that the users may change their answers after the first judgement. To summarize, we consider that the WM is more suitable to model the cognition in IR.

Quasi-Weak Measurement Model (QWM)

We test our model on the TREC Session Track 2013 and 2014. In IR, documents, queries can be represented in semantic space as column vectors (i.e. quantum state) by word embedding (e.g. Word2vec or Doc2vec). The dynamic user's IN can also be represented in this space.

Model	TREC 2013		TREC 2014	
	$nDCG@10$	MRR	$nDCG@10$	MRR
Win-Win (Luo, Dong, and Yang 2015)	0.1278	0.3821	0.2111	0.3859
DPL (Luo, Dong, and Yang 2015)	0.1524	0.4125	0.2301	0.4431
QMT (Wang et al. 2017)	0.1485	0.3942	0.2123	0.4402
QWM	0.1554	0.4256	0.2253	0.4475

Table 1: Evaluation Results

Model	TREC 2013		TREC 2014	
	$nDCG@10$	MRR	$nDCG@10$	MRR
Win-Win (Luo, Dong, and Yang 2015)	0.0300	0.1116	0.0236	0.0966
DPL (Luo, Dong, and Yang 2015)	0.0400	0.1255	0.0368	0.1220
QMT (Wang et al. 2017)	0.0453	0.1223	0.0455	0.1251
QWM	0.0521	0.1546	0.0462	0.1570

Table 2: Performance in Difficult Session

The weak value can describe the shift of user's IN. The larger the weak value is, the more users' IN may shift after browsing the corresponding document according to the context. Further, the document is more important to describe the change of users' IN according to the context. We regard the weak value as weights to accumulate the shifts of user's IN when browsing documents in history log, which is the past state of user's IN. The pseudo-relevant feedback documents are used to predict user's future IN. Thus we use Equation 5 to calculate the scores of alternative documents in current retrieval context and then use the scores to rank documents.

Table 1 is the evaluation results, which show the effectiveness of our model. We also choose some difficult sessions, in which the Language Model performs badly with less than nDCG@10 0.05. The result shows the effectiveness of our model when handling the difficult session logs.

Acknowledgements

This work is funded in part by the Key Project of Tianjin Natural Science Foundation (15JCZDJC31100), the Alibaba Innovation Research Foundation 2017, the Major Project of Chinese National Social Science Fund (14ZDB153) and MSCA-ITN-ETN-European Training Networks Project (QUARTZ).

References

Aharonov, Y., and Vaidman, L. 2008. The two-state vector formalism: an updated review. In *Time in quantum mechanics*. Springer. 399–447.

Luo, J.; Dong, X.; and Yang, H. 2015. Session search by direct policy learning. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, 261–270. ACM.

Wang, B.; Zhang, P.; Li, J.; Song, D.; Hou, Y.; and Shang, Z. 2016. Exploration of quantum interference in document relevance judgement discrepancy. *Entropy* 18(4):144.

Wang, P.; Hou, Y.; Li, J.; Zhang, Y.; Song, D.; and Li, W. 2017. A quasi-current representation for information needs inspired by two-state vector formalism. *Physica A: Statistical Mechanics and its Applications* 482:627–637.