Dynamic Content Monitoring and Exploration using Vector Spaces

Benyou Wang
wang@dei.unipd.it
Department of Information Engineering, University of Padova
Padova, Italy

ABSTRACT

This doctoral research project investigates using Quantum Theory (QT) to represent language, especially in some dynamic scenarios, e.g. when dealing with dynamic corpora or interactive tasks. The author plans to propose a quantum state driven framework for language problems and generalize it in a high-dimensional tensor space. Dynamics will be modeled by the formalism thereof of quantum evolution governing the update of quantum states. The author argues that this proposal will pave the way towards a new paradigm which may provide some novel insights about how to represent the language and its evolution in dynamic scenarios.

ACM Reference Format:
Benyou Wang. 2019. Dynamic Content Monitoring and Exploration using Vector Spaces. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19), July 21–25, 2019, Paris, France.* ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/3331184.3331412

1 MOTIVATION

Representing, retrieving and matching information is one of the key concerns of the Information Retrieval community. In real-world applications like dialogue systems and dynamic corpora, information may involve time spans and evolve in a dynamic process. Current state-of-art methods for representing, retrieving and matching tend to empirically adopt neural network (NN) approaches. This proposal argues that the NN solutions not only lack effective interpretation [4] and well-defined formulation in neural IR, but they also are limited regarding addressing the dynamic issues.

Inspired by the insights of Quantum IR [5, 6] and possible quantum phenomenon in language and IR [1, 8], this doctoral research project resorts to QT to reformulate the information representing and matching by means of mathematically-sound vector spaces. Intuitively, a polysemous word can naturally be represented as a superposed state while the documents can be considered as a mixed system with many words, in which the interaction between words may be implicitly encapsulated in an entangled connection like the connection between particles.

2 METHODOLOGY

This project will specifically define language representations based on the abstract vector spaces utilized in QT. At a fundamental level, this proposal aims to investigate two key Research Questions (RQ), namely: ***RQ1*** *How can language representation benefit from QT?* and ***RQ2*** *How can the dynamics of language modeled with QT?*

To the aim of answering the research questions, the doctoral research program is organized with two work-packages (WPs):

WP1 *Semantic Hilbert Space.* A general Hilbert vector space will be proposed to unify sememes, words, phrases, and documents, which will be implemented by quantum probability (especially with complex-valued representations [2]) driven NN approaches [3, 7, 9].

WP2 *Content Dynamics Monitoring.* The above Semantic Hilbert Space will be extended in a higher dimension namely a tensor space [10], in which one can interpret basic NN components in a perspective of higher-dimension space. Such an extension may help address the dynamic issues of language tasks e.g. by investigating quantum state transformation.

3 EVALUATION

Regarding the effectiveness, it is expected to achieve performance generally comparable with state-of-art models and outperform them in some specific tasks. The author is also trying to evidence this proposal in terms of interpretability. One of the secondary goals may be to provide some insights into designing NN architectures, e.g., quantitatively setting the hyper-parameters.

ACKNOWLEDGMENTS

This Ph.D. project is supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 721321.

REFERENCES

[1] Peter D Bruza, Kirsty Kitto, Douglas McEvoy, and Cathy McEvoy. 2008. Entangling words and meaning. (2008).
[2] Qiuchi Li, Sagar Uprety, Benyou Wang, and Dawei Song. 2018. Quantum-inspired complex word embedding. *arXiv preprint arXiv:1805.11351* (2018).
[3] Qiuchi Li, Benyou Wang, and Massimo Melucci. 2019. CNM: An Interpretable Complex-valued Network for Matching. *arXiv preprint arXiv:1904.05298* (2019).
[4] Zachary C Lipton. 2016. The mythos of model interpretability. *arXiv:1606.03490* (2016).
[5] Massimo Melucci. 2015. *Introduction to information retrieval and quantum mechanics.* Vol. 35. Springer.
[6] Cornelis Joost Van Rijsbergen. 2004. *The geometry of information retrieval.* Cambridge University Press.
[7] Benyou Wang, Qiuchi Li, Massimo Melucci, and Dawei Song. 2019. Semantic Hilbert Space for Text Representation Learning. *arXiv preprint arXiv:1902.09802* (2019).
[8] Benyou Wang, Peng Zhang, Jingfei Li, Dawei Song, Yuexian Hou, and Zhenguo Shang. 2016. Exploration of quantum interference in document relevance judgement discrepancy. *Entropy* 18, 4 (2016), 144.
[9] Peng Zhang, Jiabin Niu, Zhan Su, Benyou Wang, Liqun Ma, and Dawei Song. 2018. End-to-End Quantum-like Language Models with Application to Question Answering. *AAAI.* (2018), 5666–5673.
[10] Peng Zhang, Zhan Su, Lipeng Zhang, Benyou Wang, and Dawei Song. 2018. A Quantum Many-body Wave Function Inspired Language Modeling Approach. In *CIKM.* ACM, 1303–1312.

SIGIR '19, July 21–25, 2019, Paris, France

ACM ISBN 978-1-4503-6172-9/19/07.
https://doi.org/10.1145/3331184.3331412