Quantum Natural Language Processing

Bob Coecke, Giovanni de Felice, Konstantinos Meichanetzidis, Alexis Toumi
Cambridge Quantum Computing Ltd. & Oxford University, Department of Computer Science

Stefano Gogioso and Nicolo Chiappori
Hashberg

By Quantum Natural Language Processing (QNLP) we mean the canonical implementation of natural language on quantum hardware, where by canonical we mean that compositional language structure, including grammar, matches the manner in which quantum systems compose.

The Categorical Distributional Compositional (DisCoCat) model for natural language [8] enables such a canonical embedding. One instance of this is the perfect match of grammatical structure in terms of pregroups [15] and the compositional quantum structure of bipartite entanglement [1]. In fact, DisCoCat was directly inspired by teleportation-alike behaviours [5].

Besides vector spaces and inner-products, which are commonplace in modern Natural Language Processing (NLP), DisCoCat also employs several other quantum-theoretic features, such as projector spectra for representing meanings of adjectives, verbs and relative pronouns [17, 12, 13, 7], density matrices for representing linguistic ambiguity and lexical entailment [16, 2], and entanglement for representing correlated concepts [4], all of which 'exits' on quantum hardware. Therefore DisCoCat-QNLP deserves to be referred to as 'quantum-native'.

The first proposal to implement QNLP was put forward in [19]. A first major upshot of quantum implementation of DisCoCat is an exponential reduction of space resources as compared to implementations on classical hardware. Other initially mentioned upshots include the nativeness of density matrices, and the availability of quantum algorithms that provide an algorithmic quantum advantage for typical NLP tasks such as classification.

However, a 1st shortcoming in that proposal was the reliance on quantum RAM [11], which doesn't exist yet, and may never do. Also, one needs to provided hardware-dependent conversion of DisCoCat diagrams into, e.g. quantum circuits. These shortcoming are addressed in:

- A paper submitted to QPL by us:

  Quantum natural language processing on near-term quantum computers, 2020, arXiv:2005.04147.

Recently, we performed QNLP in the form of question-answering on IBM quantum hardware. In fact, this was the 1st time any form of NLP had been done on quantum hardware. The main two resources for our implementation are:

- A medium blog describing the implementation that we did:

  https://medium.com/cambridge-quantum-computing/quantum-natural-language-processing-748d6f27b31d

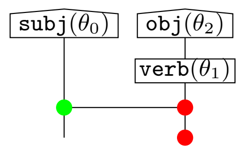
- A github repository containing the implementations:

  https://github.com/oxford-quantum-group/discopy/blob/ab2b356bd3cad1dfb55ca6606d6c4b4181fe590c/notebooks/qnlp-experiment.ipynb

This work makes use of tool box components presented in:

- Another ACT submission describing the DisCoPy toolbox in that stack [10].

One key change as compared to [19] is the use of variational quantum circuits [3] instead of qRAM. Here is such a parametrised quantum circuit that we used:



where the values of $\theta_0, \theta_1, \theta_2$ are learned using a small corpus. For reasons of simplicity, we used the verb structure that was studied in [14]. The task, rather than classification, is question answering [9], and we used a 1-dimensional sentence space. We made use of ZX-calculus [6] for easy translation between between DisCoCat diagrams and quantum circuits, and used CQC's t|ket> compiler and optimisation [18], which also relies on ZX-calculus.

Acknowledgement.

KM was supported by the EPSRC National Hub in Networked Quantum Information Technologies. AT thanks Simon Harrison for his generous support for the Harrison Quantum Foundations Fellowship he enjoys.

References

[1] S. Abramsky and B. Coecke. A categorical semantics of quantum protocols. In *Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science (LICS)*, pages 415–425, 2004. arXiv:quant-ph/0402130.

[2] D. Bankova, B. Coecke, M. Lewis, and D. Marsden. Graded entailment for compositional distributional semantics. *Journal of Language Modelling*, 6(2):225–260, 2019. arXiv:1601.04908.

[3] M. Benedetti, E. Lloyd, and S. Sack. Parameterized quantum circuits as machine learning models. *arXiv preprint arXiv:1906.07682*, 2019.

[4] J. Bolt, B. Coecke, F. Genovese, M. Lewis, D. Marsden, and R. Piedeleu. Interacting conceptual spaces I: grammatical interaction of concepts. In M. Kaipainen, A. Hautamäki, P. Gärdenfors, and F. Zenker, editors, *Concepts and their Applications*, Synthese Library, Studies in Epistemology, Logic, Methodology, and Philosophy of Science. Springer, 2018. to appear.

[5] S. Clark, B. Coecke, E. Grefenstette, S. Pulman, and M. Sadrzadeh. A quantum teleportation inspired algorithm produces sentence meaning from word meaning and grammatical structure. *Malaysian Journal of Mathematical Sciences*, 8:15–25, 2014. arXiv:1305.0556.

[6] B. Coecke and R. Duncan. Interacting quantum observables. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming (ICALP)*, Lecture Notes in Computer Science, 2008.

[7] B. Coecke, M. Lewis, and D. Marsden. Internal wiring of cartesian verbs and prepositions. In M. Lewis, B. Coecke, J. Hedges, D. Kartsaklis, and D. Marsden, editors, Procs. of the 2018 Workshop on *Compositional Approaches in Physics, NLP, and Social Sciences*, volume 283 of *Electronic Proceedings in Theoretical Computer Science*, pages 75–88, 2018.

[8] B. Coecke, M. Sadrzadeh, and S. Clark. Mathematical foundations for a compositional distributional model of meaning. In J. van Benthem, M. Moortgat, and W. Buszkowski, editors, *A Festschrift for Jim Lambek*, volume 36 of *Linguistic Analysis*, pages 345–384. 2010. arxiv:1003.4394.

[9] G. De Felice, K. Meichanetzidis, and A. Toumi. Functorial question answering. *arXiv:1905.07408*, 2019.

[10] G. de Felice, A. Toumi, and B. Coecke. DisCoPy: Monoidal categories in python. *arXiv preprint arXiv:1406.3056*, 2020.

[11] V. Giovannetti, S. Lloyd, and L. Maccone. Quantum random access memory. *Physical review letters*, 100(16):160501, 2008.

[12] E. Grefenstette and M. Sadrzadeh. Experimental support for a categorical compositional distributional model of meaning. In *The 2011 Conference on Empirical Methods on Natural Language Processing.*, pages 1394–1404, 2011. arXiv:1106.4058.

[13] D. Kartsaklis and M. Sadrzadeh. A study of entanglement in a categorical framework of natural language. In *Proceedings of the 11th Workshop on Quantum Physics and Logic (QPL)*. Kyoto Japan, 2014.

[14] D. Kartsaklis, M. Sadrzadeh, S. Pulman, and B. Coecke. Reasoning about meaning in natural language with compact closed categories and Frobenius algebras. In *Logic and Algebraic Structures in Quantum Computing and Information*. Cambridge University Press, 2015. arXiv:1401.5980.

[15] J. Lambek. From word to sentence. *Polimetrica, Milan*, 2008.

[16] R. Piedeleu, D. Kartsaklis, B. Coecke, and M. Sadrzadeh. Open system categorical quantum semantics in natural language processing. In *CALCO 2015*, 2015. arXiv:1502.00831.

[17] M. Sadrzadeh, S. Clark, and B. Coecke. The Frobenius anatomy of word meanings I: subject and object relative pronouns. *Journal of Logic and Computation*, 23:1293–1317, 2013. arXiv:1404.5278.

[18] S. Sivarajah, S. Dilkes, A. Cowtan, W. Simmons, A. Edgington, and R. Duncan. t|ket>: A retargetable compiler for NISQ devices. *arXiv preprint arXiv:2003.10611*, 2020.

[19] W. Zeng and B. Coecke. Quantum algorithms for compositional natural language processing. In *Proceedings SLPCS 2016*, volume 221 of *EPTCS*, pages 67–75, 2016.