QNLP: Compositional Models of Meaning on a Quantum Computer*

Konstantinos Meichanetzidis[†‡], **Robin Lorenz**[†], **Anna Pearson**[†], **Alexis Toumi**[†‡], **Giovanni de Felice**[†‡], **Dimitri Kartsaklis**[†], **Bob Coecke**[†]

[†]*Cambridge Quantum Computing Ltd, 17 Beaumont Street, Oxford OX1 2NA, UK*
[‡]*Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford OX1 3QD, UK*

Introduction. DisCoCat (DIStributional COmpositional CATegorical) (Coecke et al., 2010) is a framework for models of natural language meaning that comes with a rigorous treatment of the interplay between syntax and semantics and with a convenient diagrammatic representation in terms of string diagrams. The conception of this framework was the fruit of recognising the shared formal structure between *pregroup grammar* (Lambek, 2008) and compact closed categories like that of finite-dimensional Hilbert spaces (*FHilb*). A choice of model within the framework essentially amounts to choosing an appropriate semantics category and a functor into it from the (pregroup) grammar category. This functor is the mechanism through which a sentence's syntax determines, in a principled way, how the word representations interact to yield the sentence meaning – effectively, sentences are represented as string diagrams with an open wire carrying the sentence meaning interpreted in the chosen semantics category. The motivation for such a framework stems from the ambition to reconcile vector space semantics with formal approaches to linguistics, and to address the question of how the meaning of a sentence arises from the meanings of its words. Over the past 10 years DisCoCat also attracted interest as it was demonstrated to be useful for capturing linguistic phenomena such as ambiguity and entailment (Bankova et al., 2019; Kartsaklis and Sadrzadeh, 2013). Furthermore, part of the motivation for DisCoCat models is the interpretability of language models, which is a quality that does not characterise modern approaches to natural language processing (NLP).

Above all, however, as had already been noted early on, if choosing to interpret the string diagrams in *FHilb* as *quantum circuits* then the 'computation of meaning' – a tensor contraction – would most naturally be performed on a *quantum* computer. In this case the DisCoCat model makes a correspondence between words and quantum states and between grammatical structure and Bell effects. Notably, the work of Zeng and Coecke (Zeng and Coecke, 2016) built on this idea and presented a quantum algorithm for sentence similarity by reduction to the closest vector problem. Today, noisy intermediate-scale quantum (NISQ) processors are available via cloud access, naturally providing an opportunity for implementing simple NLP tasks using DisCoCat. The theoretical basis for such near-term implementations has since been laid out in the works of Refs. (Meichanetzidis et al.; Coecke et al., 2020).

Here, we report on a series of experiments, presented in Refs. (Meichanetzidis et al., 2020) and (Lorenz et al., 2021), which are the first experiments that implement a DisCoCat model – in fact any NLP model – on an actual quantum machine. All three experiments successfully address simple *binary classification* tasks.

The tasks. The first task as a basic proof of concept, addressed in Ref. (Meichanetzidis et al., 2020), takes the labels of the sentences in a very small dataset of 16 sentences as 'truth values'. The second task concerns a 'meaning classification' of sentences – whether they are about either 'food' or 'IT' – and has a dataset of 130 sentences. The third task concerns the syntactical role of 'relative pronouns' in noun phrases – whether the relative pronoun replaces the subject or the object of the respective relative subclause – and has a dataset of 105 sentences. The second and third tasks are addressed in Ref. (Lorenz et al., 2021).

The experiments. The general pipeline that is depicted in Fig. 1 shows the basic idea underlying

*Extended abstract as non-proceedings submission to *ACT* 2021 for a contribution based on the two preprints `arXiv:2102.12846` and `arXiv:2012.03756` (Refs. (Meichanetzidis et al., 2020) and (Lorenz et al., 2021) below).

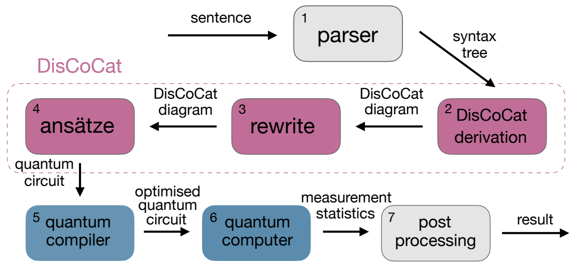


Figure 1: Schematic overview of the general pipeline.

all three experiments. It begins by parsing each sentence using a pregroup grammar built out of atomic types n for *nouns* and s for *sentence*. The name of a pregroup grammar stems from the fact that every atomic type $t \in \{n, s\}$ has a left- and a right- adjoint type (t^l and t^r), with the property that $t^l t \to 1$ and $tt^r \to 1$, where 1 is the trivial type. The existence of two different inverses is motivated by the fact that in language word-order can carry meaning. The pipeline then involves translating each type-tagged sentence into the induced DisCoCat diagram, a string diagram. This diagram represents the grammatical reduction of the sentence and can be seen as a graphical proof that the sentence is grammatical, as witnessed by the reduction of the product of types of all words to the s-type. Finally the boxes are filled with parametrised quantum circuits. Evaluating the circuits on a quantum computer returns outcome statistics from which the labels are estimated. The model is trained by updating the circuit parameters, in a supervised-learning style, to optimise the agreement between the predicted and actual labels.

Consider as an example the sentence ‘person prepares tasty dinner’ (from the second task’s dataset). Fig. 2a shows its DisCoCat diagram based on the pregroup parsing with the cups corresponding to the pregroup reductions $n \cdot n^r \to 1$ and $n^l \cdot n \to 1$ and the output type indeed being that of a sentence. Fig. 2b shows a corresponding quantum circuit as the output of step (4) of Fig. 1, where essentially three steps have happened. First, the n and s type wires were all assigned a single qubit (more generally, while beyond the capacities of the currently available quantum devices, a different number of qubits may be assigned to each pregroup type). These qubits define Hilbert spaces in which the word meanings are to be represented as quantum states. Second, all states have been assigned concrete parametrised quantum states. Specifically, these states are prepared from a trivial reference

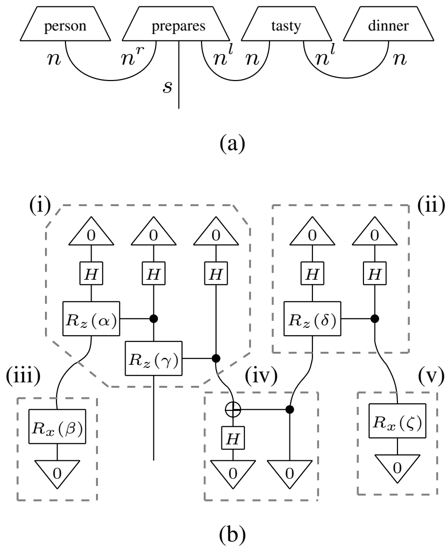


Figure 2: Example sentence ‘person prepares tasty dinner’ with in (a) its DisCoCat diagram reflecting the pregroup parsing and in (b) a parametrised quantum circuit it can be mapped to.

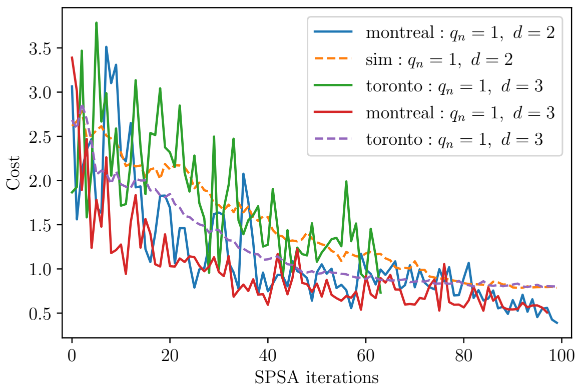
state by parameterised quantum circuits. Thus, a word state is defined by the parameters of the quantum circuit that prepares it. A choice of consistently assigning number of qubits to wires and quantum states to words is termed an *ansatz* – the step that determines the concrete parametrisation of the word embeddings. Third, the states corresponding to *person* and *dinner* were turned into effects, by ‘bending them down’, in order to reduce the overall number of needed post-selections, incurred by nondeterministic Bell effects corresponding to cups.

Having split the datasets into respective train and test subsets, the model parameters are trained on the former subset via the SPSA[§] optimisation algorithm against the cross-entropy Cost function, measuring the discrepancy between predicted and actual labels. Despite the typical noise that comes with currently available NISQ machines, in all three experiments the model converges well, i.e. the Cost is minimised successfully.[‡] In addition, classical simulations were performed to see the projected behaviour of the model in a noise-free set-up. The typical errors after 100 SPSA iterations are around 8-25% on the training data and 17-37% on the test data, depending on which of the three

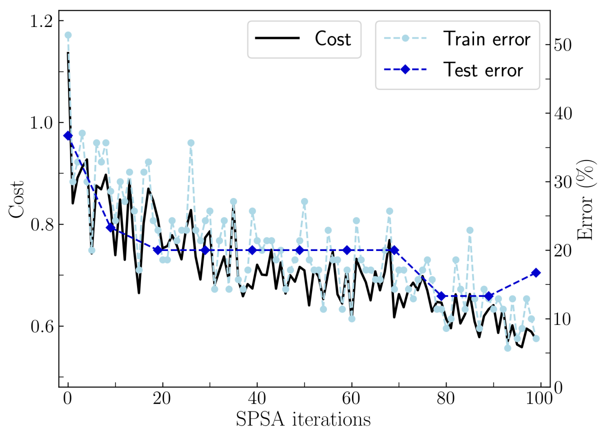
[§]Simultaneous perturbation stochastic approximation: the optimiser we used to train the model.

[‡]We used IBM’s machines `ibmq_montreal`, `ibmq_toronto` and `ibmq_bogota` with log_2QuantumVolume $= 5$.

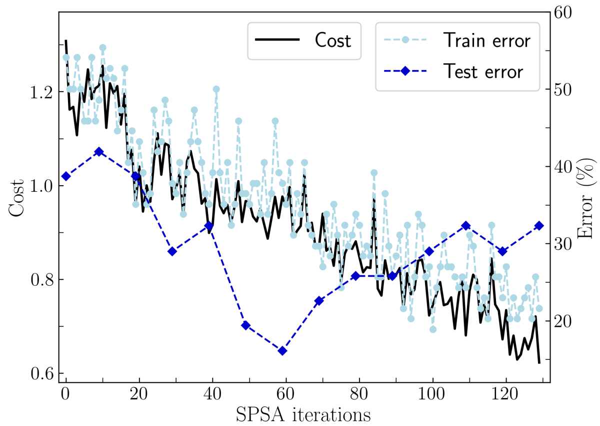
tasks.



(a)



(b)



(c)

Figure 3: Results for the first, second and third task in (a), (b) and (c), respectively (q_n and d are hyperparameters determining the ansatz). See Refs. (Meichanetzidis et al., 2020),(Lorenz et al., 2021) for details.

Conclusions. From a quantum machine learning perspective, this is an instance of a variational quantum circuit approach, where, importantly, the structure of the circuit, that is its connectivity, is not rooted in mere heuristics, but in fact dictated by the sentence's syntax. From an NLP perspective, contemplating an obvious question today, namely whether one can do NLP on a quantum computer, the work serves as proof of concept and indeed paves the way to such QNLP. It also lends support to seeing DisCoCat as a natural choice of a language model to that end. Future work may further scale up the NLP tasks one can consider as the available quantum machines improve, do comparative analyses with approaches that do *not* use a compositional model like DisCoCat, that hardwires grammar and finally, explore the scope of an experiment demonstrating a possible quantum advantage for NLP tasks.

References

Dea Bankova, Bob Coecke, Martha Lewis, and Dan Marsden. 2019. Graded Entailment for Compositional Distributional Semantics. *Journal of Language Modelling*, 6(2):225–260.

Bob Coecke, Giovanni de Felice, Konstantinos Meichanetzidis, and Alexis Toumi. 2020. Foundations for Near-Term Quantum Natural Language Processing.

Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark. 2010. Mathematical Foundations for a Compositional Distributional Model of Meaning. *Linguistic Analysis*, 36:345–384.

Dimitri Kartsaklis and Mehrnoosh Sadrzadeh. 2013. Prior Disambiguation of Word Tensors for Constructing Sentence Vectors. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNL)*, pages 1590–1601.

J. Lambek. 2008. *From Word to Sentence*. Polimetrica, Milan.

Robin Lorenz, Anna Pearson, Konstantinos Meichanetzidis, Dimitri Kartsaklis, and Bob Coecke. 2021. QNLP in Practice: Running Compositional Models of Meaning on a Quantum Computer.

Konstantinos Meichanetzidis, Stefano Gogioso, Giovanni De Felice, Nicolò Chiappori, Alexis Toumi, and Bob Coecke. Quantum natural language processing on near-term quantum computers.

Konstantinos Meichanetzidis, Alexis Toumi, Giovanni de Felice, and Bob Coecke. 2020. Grammar-Aware Question-Answering on Quantum Computers.

William Zeng and Bob Coecke. 2016. Quantum Algorithms for Compositional Natural Language Processing. *Electronic Proceedings in Theoretical Computer Science*, 221:67–75.