Structure-aware learning of word-embeddings using variational quantum circuits

Daphne Wang[1,2] and Mehrnoosh Sadrzadeh[1]

[1]University College London [2]Quandela

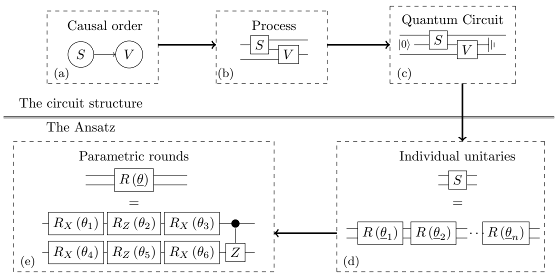
There are numerous words in English such as “spring” or “bank” which have several possible meanings, e.g. “spring” either a season or a metal coil. We say that such words are ambiguous. Several factors determine how humans decide which meaning to select, such as context and grammar, which are studied in psycholinguistics. On the other hand, much progress has been made on automatic disambiguation in Natural Language Processing (NLP), notably using deep neural networks.

In this work, we train variational quantum circuits to model the disambiguation process of ambiguous phrases (i.e. containing multiple ambiguous words). Our aim is two-fold: first, we want to make use of the inherent probabilistic nature of quantum systems to model ambiguity, and second, by using structural information, our goal is to obtain a more cognitively plausible representation of the disambiguation process than current approaches in classical machine learning. Indeed, standard practices in classical machine learning do not take grammar into account, even though it is clear from psycholinguistics that words of different grammatical types are disambiguated differently[8]. Our approach, on the other hand, is based on psycholinguistic research as well as previously obtained results on the causal structure of the disambiguation process of ambiguous phrases[12]. The outline of this abstract is as follows: we first introduce the model of previous work [12] where a set of subject-verb (SV) and verb-object (VO) phrases were studied using the framework of [3] for quantum contextuality and causality. The obtained models are then approximated using variational quantum circuits. We then show that these circuits have the potential to be used as word-embeddings in NLP tasks.

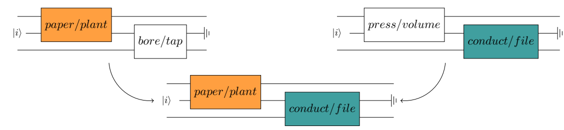
Obtaining causal orders

In [12], we studied SV and VO phrases where the subjects, objects, and verbs were all ambiguous. Each of these types of words (i.e. subject, verb, and objects) corresponded to parties or labs. Each of these parties can conduct experiments, i.e. choose an operation to conduct (which corresponds to a party’s input), and obtain an outcome. In our case, the inputs will be choices of words to fill a grammatical role, and the outcomes will be the activated interpretation of this word. Now, parties can be causally ordered, i.e. they can influence each other. For example, the choice of subject could have an impact on the choice of verb in a sentence and on the interpretation of the chosen verb, in which case we will write $S \to V$ (i.e. S can influence V).

To investigate the possible causal order between the parties S, V and O, we collected human plausibility judgments of interpretation combinations in ambiguous phrases. From these plausibility ratings, we obtained probability distributions for joint meaning activations for SV and VO phrases, which will be used in the training process described below. This dataset is described in more detail in [12], and is available on [11]. In order to identify the main causal orders at the heart of the disambiguation process, we calculated the *causal fractions*, which represent the degree to which the obtained probability distributions are compatible with a given causal order [3]. In particular, we calculated the causal fractions associated with the causal order $S \to V$, $V \to S$, $V \to O$, and $O \to V$. What we found is that the causal fractions for the causal order $S \to V$ and $O \to V$ were overwhelming dominant as opposed to the other causal orders (on average 0.89 and 0.91 for $S \to V$ and $V \to O$ respectively). The conclusion is that subjects and objects tend to be disambiguated before the verb they are attached to. This was consistent with the findings reported in psycholinguistic studies which used eye-tracking technology to obtain this conclusion[8].



(a) Obtaining the ansatz



(b) Procedure for obtaining new (untrained) circuits from trained ones.

(c) Obtaining word-embeddings

Figure 1: Summary of the approach

Training quantum models

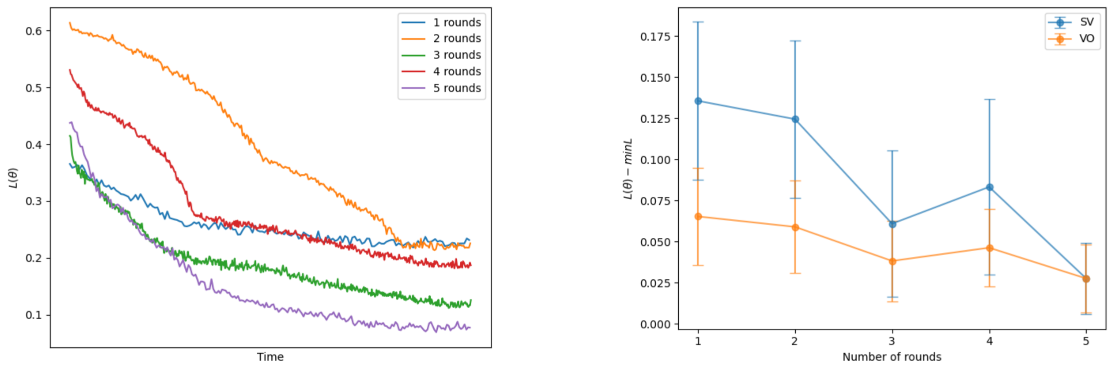
To create our quantum models, we start from the causal orders obtained in the previous section as these causal orders naturally lead to a basic structure of the process they describe[1, 4, 5], for example see Fig. 1a (a)-(b). In Fig.1a(b), the choice of inputs corresponds to the choice of words, and the outcomes represent interpretations of these words. This basic structure can then be approximated using variational quantum circuits. The parameters are trained to approximate the probability distributions coming from human data. We optimise the parameters using a fairly standard hybrid quantum-classical method [6, 2, 7, 10, 13].

The ansatz is designed as follows: firstly, we choose the input and output systems to be qubits, so that the two choices of words could be represented as $|0\rangle$ and $|1\rangle$, and similarly for their two possible meanings. Then, we want to feed a subsystem of the output of the subject/object operation to the verb operation; we take this subsystem to be $\mathbb{C}^2$ as well. Now, since the output of the nouns operations are 2-qubit systems, to satisfy the unitary condition we also need their inputs to be a 2-qubit system; we then decide to add an ancilla qubit set to $|0\rangle$ in addition to the input corresponding to the choice of noun. Similarly, there will be an extra qubit in the output of the verb-circuit, which we will discard. The form of these circuits is illustrated in Fig. 1a(c). Each individual operation is encoded as one parametric quantum circuit. The circuits are divided into rounds of single-qubit unitaries and entangling gates; see Fig. 1a(d-e). Increasing the number of rounds (and therefore the number of parameters) is expected to increase the accuracy of the circuit but will also take longer to be trained. We will choose to have the same number of rounds for both parties. Each round is depicted in Fig. 1a(e). Each qubit is subject to a X-rotation, a Z-rotation, and then another X-rotation. We apply a controlled-Z gate to create entanglement.

To train the parameters, we use Qiskit's Aer platform [9] and apply the gradient descent algorithm with finite approximation. Our cost function is the total variation between the obtained probability distribution (estimated from the counts) and the probability distribution obtained from the human judgment dataset. We trained our circuits for numbers of rounds varying from 1 to 5 and a random seed. We found that all of the models converged, see Fig. 2a. Furthermore, the converged cost gets closer to the minimal possible cost (which can be calculated from the causal fraction) as the number of rounds increases; see Fig.2b. This shows that the accuracy of the quantum circuits increases as the number of parameters increases. We also observed that the choice of initial parameters had very little impact on the final cost of the models. However, the values of the optimised parameters were highly dependent on the initial parameters. Therefore, we fix the choice of initial parameters in the subsequent results to reduce the parameters' variability.

Predicting meaning combinations

We then constructed a set of models which we trained using the procedure described above; these models constitute our training set. We then test the predictions obtained from recombining the subject, object, and



(a) Cost function for an SV model.

(b) Average optimised cost (offset by minimal costs).

Figure 2: Convergence of the variational circuits

verb circuits (as in Fig. 1b) to predict the probability distributions of new SV and VO models which we can compare with human data; these new empirical models will constitute our testing set. We observe that the predicted circuits achieve a reasonably high accuracy, namely 0.76 for the SV models and 0.86 for the VO models. This shows that our predicted models are indeed learning transferable information.

Obtaining quantum word embeddings

A natural question that arose was whether our trained circuits lead to meaningful representations of words and could thus be used as quantum embeddings for words. If so, they should be usable in mainstream NLP tasks such as word-sense disambiguation. We here test the viability of this hypothesis, and leave testing the obtained embeddings in NLP tasks to future work.

Each word-state is trained for a specific model. Hence, we first check that the states which should correspond to the same word are indeed similar. If this were not the case, these word-states would not be useful anyway as a single word would have multiple representations. Here, we will assume that each word's dependency constitutes an inherent part of the word. For example, the noun *pitcher* used as a subject will be considered distinct from the same word used as an object. Using circuits trained as described above, we obtain a quantum state representation of a subject or object by fixing the input state (as opposed to the ancilla state which is constant) of the S or O individual circuits (see Fig. 1c). For verbs, the process is a little more complicated as the representation of the verb not only depends on the choice of the verb but also on the outputs of S or O. Hence, to obtain the verb representation, we first fix its input and then take the *partial trace* over the subsystem dependent on the S or O output (see Fig. 1c). This procedure will give us a density matrix instead of a pure quantum state. The similarity between words will then be quantified by their inner-products.

We found out that the states corresponding to the same nouns have a large overlap across models, with an average inner product of 0.64. For verb-states, the overlap between verb-states from different models was lower and on average 0.37. On the other hand, by restricting the representations of a word that correspond to the same input state (i.e. $|0\rangle$ or $|1\rangle$), we are able to boost the inner products to 0.94 for noun-states and 0.48 for verb-states. Overall, this means that the quantum word embeddings that we obtain from these variational circuits have the potential to be useful in NLP tasks as long as one fixes the input state it corresponds to.

Summary

In this work, we trained variational quantum circuits for predicting the meaning combinations of ambiguous two-word phrases. The structure of these circuits was fixed by previous work and psycholinguistic theories. We then managed to use the optimised circuits to predict the meaning activation patterns of unseen phrases and obtain meaningful representations of ambiguous words. This work therefore offers great potential for cognitively plausible quantum natural language processing.

References

[1] T. Eggeling, D. Schlingemann, and R. F. Werner. Semicausal operations are semilocalizable. *Europhysics Letters*, 57(6):782, mar 2002.

[2] Edward Farhi and Hartmut Neven. Classification with Quantum Neural Networks on Near Term Processors, 2018.

[3] Stefano Gogioso and Nicola Pinzani. The Sheaf-Theoretic Structure of Definite Causality. *Electronic Proceedings in Theoretical Computer Science*, 343:301–324, Sep 2021.

[4] Aleks Kissinger, Matty Hoban, and Bob Coecke. Equivalence of relativistic causal structure and process terminality, 2017.

[5] Aleks Kissinger and Sander Uijlen. A categorical semantics for causal structure. *Logical Methods in Computer Science*, 15, 2019.

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

[7] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii. Quantum circuit learning. *Phys. Rev. A*, 98:032309, Sep 2018.

[8] Martin J. Pickering and Steven Frisson. Processing ambiguous verbs: Evidence from eye movements. *Journal of Experimental Psychology: Learning Memory and Cognition*, 27, 2001.

[9] Qiskit contributors. Qiskit: An open-source framework for quantum computing, 2023.

[10] Maria Schuld, Alex Bocharov, Krysta M. Svore, and Nathan Wiebe. Circuit-centric quantum classifiers. *Physical Review A*, 101(3), mar 2020.

[11] Daphne Wang. `https://github.com/wangdaphne/Causality-of-lexically-ambiguous-phrases`, 2022.

[12] Daphne Wang and Mehrnoosh Sadrzadeh. The causal structure of semantic ambiguities, 2023.

[13] D. Zhu, N. M. Linke, M. Benedetti, K. A. Landsman, N. H. Nguyen, C. H. Alderete, A. Perdomo-Ortiz, N. Korda, A. Garfoot, C. Brecque, L. Egan, O. Perdomo, and C. Monroe. Training of quantum circuits on a hybrid quantum computer. *Science Advances*, 5(10):eaaw9918, 2019.