Quantum Deep Sets and Sequences

Vladimir Vargas-Calderón*

Zapata AI, 100 Federal St., 02110, Boston, MA, USA

MOTIVATION

Machine learning (ML) is usually concerned with modelling functions that look like $f : \mathbb{R}^d \to \mathcal{Y}$ in the case of supervised learning (such as classification and regression) and some unsupervised learning algorithms (such as clustering or generative modelling). This paradigm restricts the kind of tasks that can be learnt using these ML algorithms. Examples of important tasks not treatable with these ML models are point cloud classification and regression [1], galaxy red-shift estimation [2], set anomaly detection [2], text concept set retrieval [3], image tagging [4], among others. The common factor between these tasks is that they can be posed as learning a function in a set space, instead of the customary vector space.

Zaheer *et al.* [2] introduced the Deep Sets model to learn functions take the form $f(X) = h\left(\sum_{x\in X} g(x)\right)$, which are functions that map a set X of objects x to a given codomain $\mathcal{Y}$. This map is done through the action of an embedding function g and a projection function h. The embedding function g takes the objects x into an embedding space; then, the sum over the embedded objects (which ensures permutation invariance) creates a new object in the embedding space containing information from all the objects in the original set X; finally, the projection function h maps the resulting embedding of the set into the specified codomain $\mathcal{Y}$. Deep Sets leverage neural networks to parameterise the functions h and g, creating a powerful framework for processing sets. The embedding function g plays a crucial role by encoding complex data correlations into a high-dimensional space. When the vectors in this space are summed, the resultant vector effectively summarises the set's information, irrespective of the set's size.

Quantum physical systems are particularly advantageous for this task. They inherently display and exploit intricate correlations, allowing them to naturally encode complex probability distributions. This makes them well-suited for enhancing the performance of Deep Sets, especially in scenarios involving rich and complex data interrelationships. This work focuses on generalising the definition of Deep Sets to accommodate quantum systems at its core, namely, by making the embedding space be that of a quantum system.

METHODS AND RESULTS

We can more generally write functions of sets by the expression $f(X) = h\left(\bigcirc_{x\in X} g(x)\right)$, where $\circ$ is any binary operation, which, in principle, should be commutative and associative to keep the variadic and permutation-invariant properties. In this work, g is a function that maps data to the quantum state of a system, i.e., g is a quantum feature map. An example of a binary operation that is commutative and associative on normalised quantum states is convolution albeit not being physically realisable in the general case [5]. Thus, we also consider the most general possible expression for functions of sets which can be given by $f(X) = h(v(\{g(x)\}_{x\in X}))$, where v is a variadic operation acting on a set.

State vector averaging.—Consider an embedding function g implemented through a variational circuit $g(x, \boldsymbol{\phi})$, which maps data x to the state vector of an n-qubit quantum register. Let us also use state vector averaging as the variadic operation acting on sets of quantum states $\{|\psi\rangle_i\}$ on the Hilbert space

* vladimir.vargas@zapata.ai

of an n-qubit system:

$$v : \{|\psi\rangle_i\} \mapsto v(\{|\psi\rangle_i\}) = \frac{\sum_i |\psi\rangle_i}{\|\sum_i |\psi\rangle_i\|}, \tag{1}$$

which is permutation invariant. These components enable the modeling of functions of sets. As a particular example, let us follow a learning task proposed by the Deep Sets paper authors [2]. First, a dataset is built as follows: sample D angles $\{\alpha_i\}_{i=1}^{D}$ in $[0, \pi]$; for each angle, create the Gaussian distribution with mean zero and covariance matrix $R(\alpha_i)\Sigma R^T(\alpha_i)$, where Σ is a fixed random 2×2 covariance matrix; from each of these 2D Gaussians, sample between 300 and 500 2D vectors to form a set $\mathcal{M}_i = \{\vec{x}_j\}$; finally, analytically compute the entropy S_i of the first dimension of the 2D Gaussian to form a dataset of pairs $\{(\mathcal{M}_i, S_i)\}_{i=1}^{D}$. The task is: given a set of 2D vectors $\mathcal{M}$, predict its entropy S.

Zaheer *et al.* [2] used three-layer fully connected neural networks as h and g. In the present work, the embedding function is given by $g(x, \boldsymbol{\phi}) = U(x, \boldsymbol{\phi}) |0\rangle^{\otimes n}$, where U is parameterised as a $SU(2^n)$ variational quantum gate acting on n-qubits (see Ref. [6] for a definition of $SU(2^n)$ gates). The neural network h takes a quantum state of n-qubits as input, and outputs the estimation of the aforementioned entropy. Figure 1 shows the performance of regression as a function of n qubits. It is seen that the quantum version of the Deep Sets model can surpass a similar classical Deep Sets model.

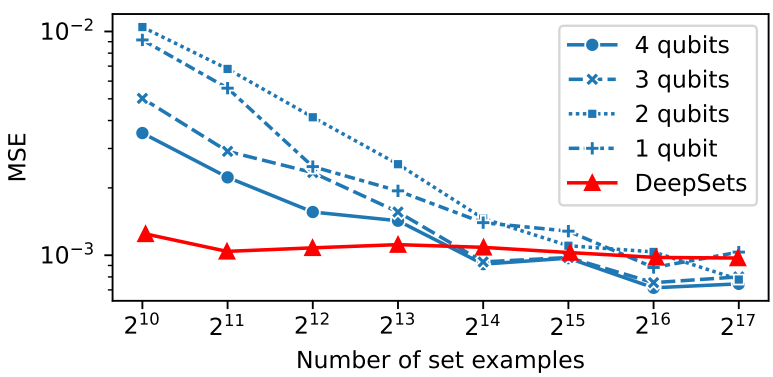

FIG. 1. Mean squared error (MSE) as a function of the number of sampled angles α (see main text). The red line corresponds to a classical Deep Sets model (data from Ref. [2]). The blue curves correspond to quantum Deep Sets models for different number of qubits n. The three-layer neural network h has 100 hidden neurons. MSE is calculated with respect to an independent set of 2^{10} angles α. Lines are for guiding the eye.

Product of density matrices.—Instead of defining variadic operations acting of quantum states such as eq. (1), consider a binary operation that can be defined with respect to a tristochastic permutation tensor T. These are three-rank binary tensors that satisfy $\sum_i T_{ijk} = \sum_j T_{ijk} = \sum_k T_{ijk} = 1$. One can define binary operations $\circ_T$ on density matrices, called binary quantum channels as [7, 8]

$$\circ_T : (\rho, \sigma) \mapsto \rho \circ_T \sigma = \mathrm{Tr}_2[U(\rho \otimes \sigma)U^\dagger], \tag{2}$$

where ρ and σ are density matrices of n qubits, Tr_2 is the partial trace over the second subsystem, and U is a particular unitary matrix (see Ref. [7] for details) that makes the binary operation in eq. (2) a binary quantum channel that is trace preserving and completely positive.

The map from a data point x to a density matrix of dimension $2^n \times 2^n$ for n qubits is given by

$$g(x, \boldsymbol{\phi}) = U(x, \boldsymbol{\phi})\, \mathrm{diag}(\boldsymbol{\lambda}(x, \boldsymbol{\phi}))U^\dagger(x, \boldsymbol{\phi}), \tag{3}$$

where U is the $SU(2^n)$ gate [6], and $\boldsymbol{\lambda} \in \mathbb{R}^{2^n}$ (see details in Ref. [9]).

In general, eq. (2) does not provide a commutative and associative binary operation [10, 11]. Thus, the order of the elements of the input set will matter and the model will be referred to as quantum Deep Sequences. In other words, this is a model for sequences of arbitrary length such as recurrent neural networks or long short-term memory (LSTM) networks. The performance of quantum Deep Sequences is showcased in the following example: create a balanced dataset of sequences of at most 50 numbers that are either sorted in increasing order or not. The task is to predict whether the sequence is sorted in increased order or not. To achieve this, the projection function h takes as input a density matrix and outputs the probability that the sequence is sorted. Figure 2 shows that the proposed quantum deep sequences model can outperform similarly sized LSTMs.

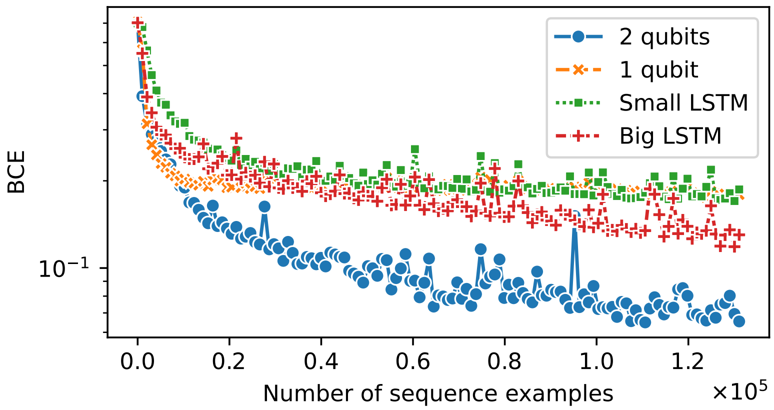

FIG. 2. Binary cross-entropy (BCE) between the true class and the predicted class as a function of the number of training examples for sorted-in-increasing-order sequence classification. The number of trainable parameters for each model are 723, 653, 2053, and 1979 for the small LSTM, 1 qubit quantum Deep Sequence, big LSTM, and 2 qubit quantum Deep Sequence, respectively. BCE is calculated with respect to an independent set of 2^{10} sequences. Lines are for guiding the eye. The orange line is mostly behind the green line.

[1] A. Goyal, H. Law, B. Liu, A. Newell, and J. Deng, Revisiting point cloud shape classification with a simple and effective baseline, in *Proceedings of the 38th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 139, edited by M. Meila and T. Zhang (PMLR, 2021) pp. 3809–3820.

[2] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov, and A. J. Smola, Deep sets, in *Advances in Neural Information Processing Systems*, Vol. 30, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017).

[3] A. H. Alsaffar, J. S. Deogun, V. V. Raghavan, and H. Sever, Concept based retrieval by minimal term sets, in *Foundations of Intelligent Systems*, edited by Z. W. Raś and A. Skowron (Springer Berlin Heidelberg, Berlin, Heidelberg, 1999) pp. 114–122.

[4] J. Fu and Y. Rui, Advances in deep learning approaches for image tagging, APSIPA Transactions on Signal and Information Processing **6**, e11 (2017).

[5] C. Lomont, Quantum convolution and quantum correlation algorithms are physically impossible (2003), arXiv:quant-ph/0309070 [quant-ph].

[6] R. Wiersema, D. Lewis, D. Wierichs, J. Carrasquilla, and N. Killoran, Here comes the SU(N): multivariate quantum gates and gradients, Quantum **8**, 1275 (2024).

[7] R. Bistroń, W. Śmiałek, and K. Życzkowski, Tristochastic operations and products of quantum states, Journal of Physics A: Mathematical and Theoretical **56**, 455301 (2023).

[8] K. Korzekwa, S. Czachórski, Z. Puchała, and K. Życzkowski, Coherifying quantum channels, New Journal of Physics **20**, 043028 (2018).

[9] K. Życzkowski, P. Horodecki, A. Sanpera, and M. Lewenstein, Volume of the set of separable states, Phys. Rev. A **58**, 883 (1998).

[10] P. Aniello, Covariant stochastic products of quantum states, Journal of Physics: Conference Series **1416**, 012002 (2019).

[11] P. Aniello, A class of stochastic products on the convex set of quantum states, Journal of Physics A: Mathematical and Theoretical **52**, 305302 (2019).