Characterizing Quantum Classifier Utility in Natural Language Processing Workflows

Kathleen Hamilton*, Mayanka Chandra Shekar, John Gounley
Computational Science and Engineering Division,
Oak Ridge National Laboratory
Oak Ridge, TN 37830 USA
Email: *hamiltonke@ornl.gov

Dhanvi Bharadwaj
Department of Physics
University of Wisconsin-Madison
Madison, WI 53706 USA

Prasanna Date
Computational Science and Mathematics Division
Oak Ridge National Laboratory
Oak Ridge, TN 37830 USA

Eduardo Antonio Coello Pérez, In-Saeng Suh, Georgia Tourassi
National Center for Computational Sciences
Oak Ridge National Laboratory
Oak Ridge, TN 37830 USA

***Abstract*—Quantum Natural Language Processing (QNLP) develops natural language processing (NLP) models for deployment on quantum computers. We explore feature and data prototype selection techniques to address challenges posed by encoding high dimensional features. Our study develops a hybrid workflow for quantum machine learning models that includes classical feature pre-processing, quantum embedding and quantum model training. To showcase a real-world application of QNLP on High-Performance Computing (HPC) and Noisy Intermediate-Scale Quantum (NISQ) devices, we employed the CANDLE/MOSSAIC datasets to compare the performance of quantum machine learning models to classical shallow machine learning methods on binary and multi-class classification tasks. We observe comparable performance in terms of recall and accuracy between the quantum and classical models, even with large datasets. These results provide a point of comparison between quantum and classical models on real-world datasets.**

***Index Terms*—quantum natural language processing, quantum neural networks, quantum machine learning**

Introduction QNLP aims to develop large language models that leverage the inherent properties of quantum mechanics to tackle complex linguistic problems on quantum computers [1]. Quantum models offer a potential solution as their operation on quantum computers allows for the exploration of exponentially larger solution spaces, which can potentially lead to more accurate results in a shorter amount of time [2], [3]. As near-term quantum hardware is advancing, there is a need to finding efficient feature embeddings that can use the increased number of qubits that can be utilized, and the increased complexity of a quantum circuit (i.e. depth). In this work we focus on the efficiency of data re-uploading with dense angle encodings.

Datasets In this work we use binary and multi-class classification of synthetic datasets to evaluate the utility of quantum classifiers. If a dataset contains unbalanced classes, then individual training samples x_i can have an associated weight c_i that gives the relative importance of correctly labeling the sample – this addresses the effects of class imbalance by making the minority class have a larger impact on the overall loss function. The relative weights are determined using the balanced heuristic weighting implemented in `scikit-learn` which is based on the methods found in [4].

The first dataset utilized for this project is the the MADELON dataset, which serves as a valuable benchmark for evaluating the efficacy of quantum machine learning algorithms in handling multi-dimensional and highly non-linear datasets. It consists of data points clustered on the vertices of an *n*-dimensional hypercube. Each data point has 20 total features of which *n* are informative, 2 are linear combinations of the informative features, and the remaining are uninformative. The data does not have attribute information to avoid biasing feature selection and 10% of the samples are randomly labeled.[1]

The second dataset employed for this study consists of a corpus of synthetically generated cancer pathology reports (CPRs). These clinical text documents describe the analysis of a tumor biopsy and are labeled for four cancer phenotyping tasks. Automating CPR classification with deep learning is important for achieving near-real-time cancer surveillance [6]. [2] P3B3 dataset includes four information extraction tasks namely site, laterality, histology and grade. For this study, we use "site" as the information extraction task.

Both datasets use classical mutual information to reduce the number of features passed to a QNN. The Madelon data used

[1]Data set generated using functionality available in `sklearn` and additional details are provided in [5].

[2]Data set details are provided in [7] and is available at https://ftp.mcs.anl.gov/pub/candle/public/benchmarks/Pilot3.

Mutual Information (MI) feature selection to extract the top 2 and 4 features from the original 20-feature dataset. This is a non-parametric method based on entropy estimation from k-nearest neighbors (k-NN) distances [8]. QNNs were trained using these extracted features and compared with the full 20-feature dataset to understand the impact of non-informative features on QNN performance.

The P3B3 dataset includes an enumerated and padded token list of sequence length 1500 for every CPRs. The dataset is pre-processed to reduce the sequence length using normalized pointwise MI [9]. This is used to rank the tokens based on the probability of occurrence of the given token by the total number of training documents [10], [11], and is used filter the important tokens in each pathology report based on any given document length, reducing the documents to one-dimensional vectors. Note that this assumption is unlike the typical NLP approach such as `word2vec` [12]. The normalized point wise MI is calculated between a token x and a *site* label y.

Quantum Neural Networks Quantum neural networks (QNNs) are parameterized quantum circuits that combine the principles of quantum mechanics and artificial neural networks to perform tasks such as pattern recognition, optimization, and classification [13], [14]. Our QNNs are constructed using parameterized two-qubit unitaries which incorporates the data encoding through data re-uploading [15]. The label prediction is obtained by projecting m qubits of the final quantum state onto a fixed basis. With $m = 1$ qubits we can assign binary labels from the probability of observing the `0` or `1` bitstring. We predict multi-class labels using $m > 1$ qubits and one-hot encoding – from the 2^m unique bitstrings, we down-select on the weight-1 bitstrings and renormalize these extracted probability amplitudes.

The choice of gates used for data re-uploading is derived from three-gate decompositions of Euler rotations: $\mathcal{D}_1 = R_X(\theta_1)R_Y(\theta_2)R_X(\theta_3)$, $\mathcal{D}_2 = R_X(\theta_1)R_Z(\theta_2)R_X(\theta_3)$ and $\mathcal{D}_3 = R_Y(\theta_1)R_Z(\theta_2)R_Y(\theta_3)$ [16]. These sequences use Pauli rotation gates $R_i(x) = e^{-ix\hat{\sigma}_i}$ and the angles θ_1, θ_2 embed features x_i re-scaled to $[0, 2\pi]$, and θ_3 is trainable. With these three-gate decompositions, and a n qubit circuit we can embed $4(n-1)$ unique features.

The general ansatz is built using p layers of 2-qubit unitaries which combine the decompositions $\mathcal{D}_i$, either parameterized ZZ coupling gates or using a decomposition of SU(4) operations [17]. The combination of $\mathcal{D}_1, \mathcal{D}_2$ and parameterized ZZ couplings contains $3(n-1)p$ and the trainable model is a transverse field Ising model. The combination of $\mathcal{D}_3$ and the SU(4) decomposition has $11(n-1)p$ trainable parameters. Each QNN is constructed and trained in Pennylane using supervised learning with categorical cross entropy loss functions and batch gradient descent with the Adam optimizer using batches of 32 training samples.

The QNN performance is compared to a classical convolution neural network model (CNN) that replicates a setup used in a previous study [7], but modified to take the angle embedding prepared for VQC model as the input. The CNN framework is built and evaluated using `Pytorch`. The binary classifier in the paper had around 9,900 trainable parameters and the multiclass classifer has around 10,800 trainable parameters. In comparison the largest QNN trained on P3B3 had at most $3(n-1)p = 9 \times 12 = 108$ gate parameters, the largest QNN trained on MADELON had at most $11(n-1)p = 55 \times 10 = 550$ parameters.

Results When both CNN and QNN models are given the the same (truncated, rescaled) P3B3 data we observe that the performance of a classical CNN is comparable to the performance of the QNNs, with the caveat that the data pre-processing is optimized for the QNNs and there are no guarantees that this feature formatting is ideal or optimal for the classical CNN. Previous studies and usage of classical CNNs have relied on `word2vec` embedding, where each feature input is a matrix (not a vector). In the current study is that the vector-formatted features are expanded into a matrix and then passed to the CNN. Future work will investigate the converse – use the matrix expansion of `word2vec` embeddings which are optimized for classical CNNs, and convert those into quantum circuit parameters, or in general explore the influence of longer feature vector lengths. Additionally, for both QNN and CNN, high accuracy on multi-class classification with unbalanced data remains a challenge.

Acknowledgment This research used resources of the Oak Ridge Leadership Computing Facility (OLCF) and the Compute and Data Environment for Science (CADES) at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. The research was supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.

REFERENCES

[1] B. Coecke, G. de Felice, K. Meichanetzidis, and A. Toumi. Foundations for near-term quantum natural language processing. *arXiv:2012.03755*, 2020.

[2] Robin Lorenz, Anna Pearson, Konstantinos Meichanetzidis, Dimitri Kartsaklis, and Bob Coecke. QNLP in practice: Running compositional models of meaning on a quantum computer. *arXiv preprint arXiv:2102.12846*, 2021.

[3] R. Guarasci, G. De Pietro, and M. Esposito. Quantum natural language processing: Challenges and opportunities. *Appl. Sci.*, 12:5651, 2022.

[4] Gary King and Lanche Zeng. Logistic regression in rare events data. *Political Analysis*, 9(2):137–163, 2001.

[5] Isabelle Guyon. Madelon. UCI Machine Learning Repository, 2008. DOI: https://doi.org/10.24432/C5602H.

[6] Tanmoy Bhattacharya, Thomas Brettin, James H Doroshow, Yvonne A Evrard, Emily J Greenspan, Amy L Gryshuk, Thuc T Hoang, Carolyn B Vea Lauzon, Dwight Nissley, Lynne Penberthy, et al. AI meets exascale computing: Advancing cancer research with large-scale high performance computing. *Frontiers in Oncology*, 9:984, 2019.

[7] Hong-Jun Yoon, John Gounley, M Todd Young, and Georgia Tourassi. Information extraction from cancer pathology reports with graph convolution networks for natural language texts. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 4561–4564. IEEE, 2019.

[8] Brian C. Ross. Mutual information between discrete and continuous data sets. *PLOS ONE*, 9(2), 02 2014.

[9] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical Review E*, 69(6), jun 2004.

[10] Gerlof Bouma. Normalized (pointwise) mutual information in collocation extraction. *Proceedings of GSCL*, 30:31–40, 2009.

[11] Andrew E Blanchard, Shang Gao, Hong-Jun Yoon, J Blair Christian, Eric B Durbin, Xiao-Cheng Wu, Antoinette Stroup, Jennifer Doherty, Stephen M Schwartz, Charles Wiggins, et al. A keyword-enhanced approach to handle class imbalance in clinical text classification. *IEEE journal of biomedical and health informatics*, 26(6):2796–2803, 2022.

[12] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[13] Maria Schuld, Alex Bocharov, Krysta M Svore, and Nathan Wiebe. Circuit-centric quantum classifiers. *Physical Review A*, 101(3):032308, 2020.

[14] Vojtěch Havlíček, Antonio D Córcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M Chow, and Jay M Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209, 2019.

[15] Adrián Pérez-Salinas, Alba Cervera-Lierta, Elies Gil-Fuster, and José I Latorre. Data re-uploading for a universal quantum classifier. *Quantum*, 4:226, 2020.

[16] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.

[17] Jin-Guo Liu, Liang Mao, Pan Zhang, and Lei Wang. Solving quantum statistical mechanics with variational autoregressive networks and quantum circuits. *Machine Learning: Science and Technology*, 2(2):025011, 2021.