Quantum Self Attention on Pubmed

Sabdha Ambati[1]

[1]**sa2975@njit.edu**

ABSTRACT

This paper presents the application of Quantum Self Attention Neural Network to Pubmed dataset of research papers. The model performs binary classification of the papers between Diabetes Mellitus, Experimental and Diabetes Mellitus Type 1. The model's performance is compared against Naive Bayes and Classical Self Attention model.The accuracy achieved by QSANN is 73% , suggesting it's potential in research papers classification.

Keywords: Text Classification, Quantum Self-Attention, NLP

INTRODUCTION

Natural Language Processing is an important branch of AI that gives computers the ability to understand human language. Many NLP implementations include information extraction, translation, sentiment analysis, text classification, and text generation. In text classification, we assign a category or a label to text based on its content. Research databases have millions of papers, and classifying them into different labels helps researchers find them faster.

For text classification, the Quantum Self Attention Neural Network(5), can capture more complex relationships in data with significantly fewer parameters when compared with the Classical Self Attention Neural Network. After many years of experimental research, we are currently at a point where quantum processors can demonstrate utility.(IBM)Currently quantum computers cannot outperform classical computers yet, but for specific complex tasks there is a realistic possibility that quantum computers can compete with classical computing models. To explore the utility of quantum computers, the task we considered is the text classification of PubMed dataset(2) using a hybrid model QSANN(5).

This model defines Quantum Self-Attention Layer that uses Parametric quantum circuits. It embeds input features using rotation gates and entangling layers, and uses Pauli-Z gates to measure outputs. The attention score is computed using a Gaussian function over the distance between Query and Key.

PRELIMINARIES AND NOTATIONS

- **Quantum Basics**(Gharibyan):

  1. **Qubits**: In quantum computing, information is usually represented by n-qubit quantum states over the Hilbert space $\mathbb{C}^2$. A quantum state can be represented using a unit vector $|\psi\rangle$ (the *bra-ket* notation is used to describe the column vector (ket) $|\psi\rangle$ and its complex conjugate transpose (bra) $\langle\psi|$). Applying a gate to a state can be described in bra-ket notation as:

     $$|\psi'\rangle = U|\psi\rangle$$

     where U is a unitary operator representing the quantum gate.
  2. **Quantum Gates**:
     - **RX Gate**: A rotation around the X-axis by angle θ:

       $$RX(\theta) = \exp\left(-i\frac{\theta}{2}X\right) = \begin{bmatrix} \cos\left(\frac{\theta}{2}\right) & -i\sin\left(\frac{\theta}{2}\right) \\ -i\sin\left(\frac{\theta}{2}\right) & \cos\left(\frac{\theta}{2}\right) \end{bmatrix}$$

     - **RY Gate**: A rotation around the Y-axis by angle θ:

       $$RY(\theta) = \exp\left(-i\frac{\theta}{2}Y\right) = \begin{bmatrix} \cos\left(\frac{\theta}{2}\right) & -\sin\left(\frac{\theta}{2}\right) \\ \sin\left(\frac{\theta}{2}\right) & \cos\left(\frac{\theta}{2}\right) \end{bmatrix}$$

- **CNOT Gate** (Controlled-NOT): A two-qubit gate:

$$\text{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

- **Pauli-Z Gate**:

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

- **Hadamard Gate**:

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

• **Model Architecture**:

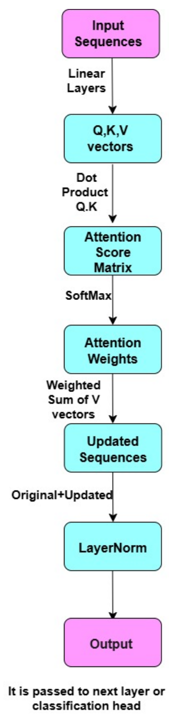

Figure 1. Classical Attention Model

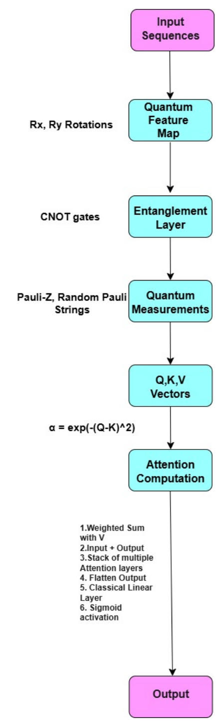

Figure 2. Quantum Self Attention Model

PROPOSED IDEA

The proposed idea is to evaluate the performance of the Quantum Self-Attention Neural Network (QSANN) against a simple Classical Self-Attention model and a Naive Bayes model on the Pubmed dataset(2).

PREPROCESSING

TF-IDF Vector Processing:
Preprocessing involved taking the TF-IDF vectors (scaled to have values between 0-100) and convert them back to pseudo-text. For words with higher TF-IDF scores, they would appear more in the text and words with very low threshold would disappear.

Missing Value Check:
The next step was to check for missing values.

Class Imbalance Handling:
For binary classification there was a class imbalance, since QSANN is built only using two qubits, down sampling of majority class was performed.

Tokenization and Padding:
With the vocabulary size as 1000 the pseudo-texts were tokenized. To maintain uniformity in the length of the texts, padding is done.

Dataset Reduction for QSANN:
For the QSANN model, due to computational resource limitations, the dataset is reduced to 128 samples in both training and test sets.

Normalization and Reshaping:
Since QSANN performs rotations R_X and R_Y on the features, the training and testing data is normalized and scaled to have values between $[-\pi/2, \pi/2]$. Next, it is reshaped to 10 rows and 20 columns, which was a 200 long vector.

Tensor Conversion:
Now the x_train set is list of torch tensors and y_train is list of tensor labels. The same steps are performed on the test dataset.

Fair Evaluation Setup:
To perform a fair evaluation, the Classical Self-Attention and Naive Bayes have the same dataset size. For Naive Bayes, the train and test sets are not normalized.

TF-IDF Baseline Performance:
The original dataset had TF-IDF values for each research paper. The accuracy when using TF-IDF was lower than 50%.

MODEL DETAILS

Classical Naive Bayes Model:
The total number of parameters were 401.

Classical Self-Attention Model:
The model had the following configuration:

- Sequence length = 10
- Number of features per sequence = 2 (analogous to number of qubits)
- Number of attention layers = 1
- Number of attention heads = 1

Total trainable parameters are 8185.

Quantum Self-Attention Model:
The model had the following configuration:

- Sequence length = 25
- Number of qubits = 2
- Number of layers for encoding = 8
- Number of variational layers = 2
- Number of QSANN layers = 1

The number of trainable parameters are 381.

The quantum gates used in this model are run on classical simulations which mimic the quantum behavior of qubits. Since the model is using `EstimatorQNN`, which internally relies on floating-point arithmetic and circuit transformations, there is a fluctuation in test and training accuracies.

TRAINING DETAILS

Quantum Self-Attention Neural Network (QSANN(qsa)):

- Loss Function: Binary Cross Entropy Loss
- Optimizer: Adam Optimizer with a learning rate of 0.005
- Batch Size: 32

- Training Procedure:
  - After each epoch, training loss and accuracy are calculated.
  - Validation loss and accuracy are monitored.
  - If validation accuracy does not improve, the learning rate is adjusted using a scheduler.
  - Early stopping is performed if the validation accuracy does not improve for 10 consecutive epochs.

Classical Self-Attention Neural Network (CSANN):

- Loss Function: Binary Cross Entropy Loss
- Optimizer: Adam Optimizer with a learning rate of 0.005
- Training Procedure:
  - Training is similar to QSANN.
  - The learning rate scheduler monitors validation accuracy.
  - Early stopping is performed if validation accuracy does not improve for 5 consecutive epochs.

Naive Bayes (NB):

- Model: Multinomial Naive Bayes (MultinomialNB)
- Hyperparameter Tuning:
  - Grid Search is used to search for the best α value.
  - 5-fold cross-validation is performed.

RESULTS

Performance Metrics

Model	Accuracy (%)	Recall Class 0 (%)	Recall Class 1 (%)	ROC-AUC (%)
QSANN	73 ± 2	66	83	79
CSANN	66	75	58	72
NB	73	75	72	75

Table 1. Performance comparison of QSANN, CSANN, and Naive Bayes models.

Confusion Matrices

QSANN Confusion Matrix:

$$\begin{bmatrix} 42 & 22 \\ 11 & 53 \end{bmatrix}$$

CSANN Confusion Matrix:

$$\begin{bmatrix} 48 & 16 \\ 27 & 37 \end{bmatrix}$$

Naive Bayes Confusion Matrix:

$$\begin{bmatrix} 48 & 16 \\ 18 & 46 \end{bmatrix}$$

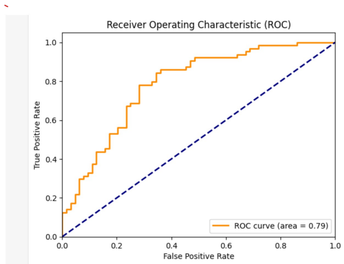

Figure 3. ROC Curve for QSANN

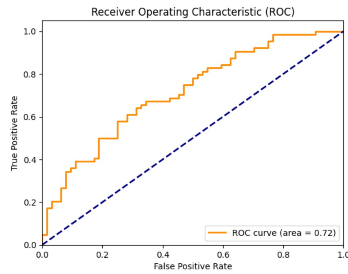

Figure 4. ROC Curve for CSANN

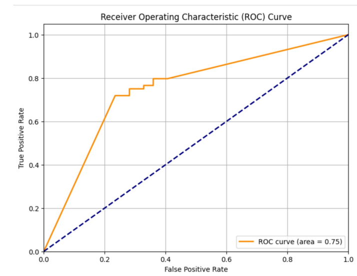

Figure 5. ROC Curve for Naive Bayes

ROC Curves

DISCUSSION OF THE RESULTS

Since the dataset is balanced, models should ideally have similar recall for both classes but QSANN is biased towards class 1, CSANN is biased towards class 0 and Naive Bayes has balanced recalls. Due to the use of EstimatorQNN, which simulates the behavior of qubits and involves random initializations, the test accuracies for QSANN tend to fluctuate. QSANN required fewer parameters compared to CSANN and Naive Bayes (NB). Due to computational limitations, the training was conducted using only two qubits and a relatively small dataset. With increasing the training dataset the peroformance can increase. NB and CSANN models were significantly faster in training and inference compared to QSANN.

REFERENCES

[qsa] @miscQSANN. https://github.com/QAMPspring2023/qgpt-issue-31. Accessed: May 5, 2025.

[2] Galileo Mark Namata, Ben London, L. G. and Huang, B. (2012). Query-driven active surveying for collective classification. In *International Workshop on Mining and Learning with Graphs*, Edinburgh, Scotland.

[Gharibyan] Gharibyan, H. @miscBreaking Down the Barriers: Quantum Computing Basics Explained. https://www.bluequbit.io/quantum-computing-basics. Accessed: May 5, 2025.

[IBM] IBM. Learning quantum ibm. https://learning.quantum.ibm.com/. Accessed: May 5, 2025.

[5] Li, G., Zhao, X., and Wang, X. (2023). Quantum self-attention neural networks for text classification.