International Journal of
Advanced Science and Convergence

ISSN : 2619-8150
Volume 3, Number 2, June 2021
https://doi.org/10.22662/IJASC.2021.3.2.007

Text Comprehension with Parameterized Quantum System

Y. Bharadwaj[1] and Dr. K. Bhanu Prakash[2]

[1]Masters Student, Department of Computer Science, K L University, Vaddeswaram, Andhra Pradesh, India
[2]Professor, Department of Computer Science, K L University, Vaddeswaram, Andhra Pradesh, India

Abstract

Background/Objectives: Information is crucial in present world; text is one form of information that is being exchanged in alarming rates. Natural language Processing is one field that concentrates on text analysis. **Methods/Statistical analysis:** Text Analyzers collects the word vectors and embed them into one by calculating semantics, and their relationships were considered on bases of dependencies and dependency trees which only targets subject to object relations and vice versa. In this digital era, microblogs involve more complicated text which are very hard using dependencies and relations to comprehend in bases of contextual semantics. **Findings:** In this paper we are addressing this problem by building a novel quantum enhanced modal. The proposed methodology exchange parameters between NLP algorithm and Quantum native optimizer allowing us to solve non-linear problems while composing the semantics. **Improvements/Applications:** We have integrated our methodology into a simple question and answering system for assessment, this system will give us the scores and answers build upon context already existing on the internet. In every Assessment Quantum Trained or Q-Trained algorithm exhibited promising results when compared with the best-in-class NLP algorithm ALBERT.

Index Terms Machine Learning, Natural language Processing, Quantum Circuit, Quantum Computing, Text processing.

Corresponding author : K. Bhanu Prakash
195036001@kluniversity.in

• Manuscript received May 18, 2021.
• Revised June 14, 2021 ; Accepted June 20, 2021.
• Date of publication June 30, 2021.

I. Introduction

Quantum computing is a branch of computer science that harness laws from quantum mechanics. For information processing, a quantum computer utilizes the mathematics behind computer science algorithms and intersect them with physics[1]. Most of the phenomenon required to understand some concepts of quantum computing deals with advanced mathematics like pre-calculus and discrete structuring concepts. However, it is easier to relate for those who are already familiar with classical computer science concepts.

Every day, computers are getting tinier but faster because of the advancements in electronic industry and because of this reduction of size, electrons may behave like transistors by transferring charge to other side themselves thus resulting in a potential quantum tunnelling. Laws and Concepts of Quantum Mechanics gives potential to solve problems in classical computation techniques. Quantum Computers uses their own structures called 'Qubits' to store information[1]. A quantum computer uses subatomic particles to represent qubits along with spins and states to represent information[2]. Model made from these could disobey the Church-Turing principle by performing calculations exponentially faster compared to a classical modal[3].

Qubit is a basic unit to represent an information in quantum computer. Like Classical bits, qubits also exist in 2 states but additional to classical bit, the qubit can also exist in a superposition state where it can make out of both states and is the primary way of storing information[4]. However, storing it is the challenging part as any minimal external interaction will warp and erroneous which may lead to indeterminacy. Entanglement can also be responsible for a qubit to be an indeterminate state[1] Entanglement can be observed in quantum space where particles relate to one another with a strong bond between them, that they will still be connected even after separated by a long distance. The occurrence of entanglement makes quantum computers much more powerful than its classical counterparts[5].

Often, we are interested in the amount of resource used by a program/algorithm to solve a particular problem called complexity which can be referred in space and time continuum[6]. For any given problem, the amount of consumption can be measured as a function which length equals to the input of instance of that problem. For example, if we take a problem of multiplying any two n bit numbers, a computer might take up to $2n^2 + 3$ units of time, the highest order here is n^2 , we say $O(n^2)$ is the order of that algorithm to solve the problem. In terms of lower bounds then we consider the lowest order and denote using Ω notation. The Church-Turing proposed that, any machine can solve any problem if that problem can be solved in a minimal machine such as a Turing machine[3]. A Turing machine is a modal consists of finite states but with infinite memory(tape). A probabilistic Turing machine can be built out of physical components and its behaviour or the state of problem solving can be easily predicted by the laws of newton physics, building a quantum Turing machine requires an amplitude resources but simulation it in a classical machine is observed to be not so efficient when considering newton's laws.

Feynman suggested that a machine completely build out of quantum physics whose laws are evaluated from quantum mechanics can be able to simulate a probabilistic Turing machine but may not obey Church-Turing thesis[7] David-Deutsch proposed a quantum Turing machine modal and its circuit system that can solve important problems effectively which implies that quantum computation has a potential to solve hard problems that a normal Turing machine cannot solve.

II. Generalized Quantum Structures

The mathematical notation behind a classical formulation is simple linear algebraic notations such as $2n^2 + 3$ similarly, in quantum computing we involve linear algebra with *Dirac* notation. In normal notation we can see that a vector is represented with an arrow on top of the variable $\vec{a}$. In Dirac notation, the same is represented as $|a\rangle$ called *'ket'*. The transpose of ket is represented as $\langle a|$ called *'bra'* and their inner product is represented as $\langle a \mid b\rangle$ called *'bra-kets'*[4].

A basic n-qubit state is described as n length binary string, but its column vector's description has 2^n components and their superposition can be represented using Bloch sphere[4]

Bloch Sphere is a representation of a 2-state quantum level system existing in pure state in a geometrical representation. The pure state is measured as a complex superposition of vectors. In Dirac notation, the pure state can be represented as

$$|\psi\rangle = \cos\theta|0\rangle + e^{\wedge}\{i\,\phi\}\ \sin\theta|1\rangle$$

In simpler terms, $|\psi\rangle = \mathrm{a}|0\rangle + \mathrm{b}|1\rangle$ where, a and b are probabilistic amplitudes measured at 0 or 1 and sum of square of those amplitudes equals to 1. Any point on the Bloch's sphere can be a cartesian coordinate as (x, y, z) [1]

where,

$$(x, y, z) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$$

These two hyper parameters (φ, θ) are required to describe a state's measurement [5]

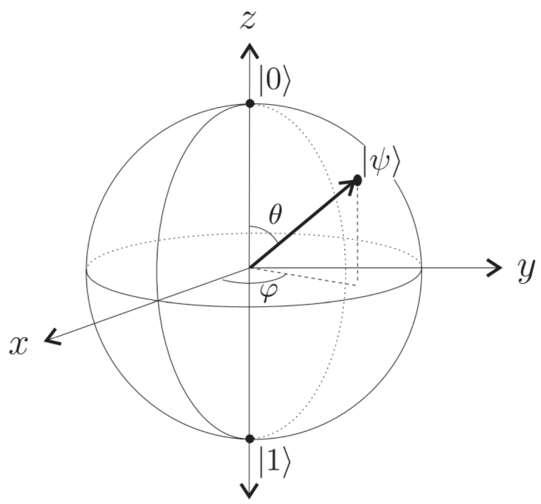


Fig 1. Bloch's sphere representing states of qubit

Since the fundamentals of quantum computing is vastly different from quantum computing methods, their properties and behaviours are also very different, some of those differences and mentioned in table 1. Qubits are not logical, they physically exist with matter, thus everything coming out of contains a direction and with anything having a direction, logical operations are not applicable however unitary operations can result in a measured value treated in a circuit system[8]

Table 1. FUNDAMENTAL DIFFERENCES BETWEEN CLASSICAL AND QUANTUM PROPERTIES [5]

Property	Classical	Quantum
Basic Unit	Binary bit	Qubit
Computing	logical	Unitary
Direction	forward	Reversable
Copying	Possible and easy	Impossible
Noise	minimal	Very difficult
Storage	Holds 1 value	Holds 2^n values

Because of their physical existence, external factors will influence the quantum operations thus causing a lot of error conditions[9]. Keeping the qubits in ideal state requires tremendous environmental changes which are highly difficult to maintain.

A single qubit itself can exhibit all the quantum principles and properties but for any problem solving a single qubit is not sufficient and will not attain any computational advantage. Like in classical computing number of bits will increase the number of possible operational outcomes as 2^n [10], qubits will take the advantage of unitary matrix giving 2^n magnitudes or amplitudes

2-bit Classical:

$$00\ 01\ 10\ 11$$

2-bit Quantum:

$$|a\rangle = a_{00}|00\rangle + a_{01}|01\rangle + a_{10}|10\rangle + a_{11}|11\rangle$$

The principles still work the same,

$$|a_{00}|^2 + |a_{01}|^2 + |a_{10}|^2 + |a_{11}|^2 = 1$$

When the states $|00\rangle$ and $|11\rangle$ are measured 50% each and $|01\rangle$ and $|10\rangle$ are measured at 0%, that state is called "Bell state" [7]

$$\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \text{ and } \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

They cannot be represented as two separate states and measuring one will immediately collapse another into the environment leaving behind a collective state, however any change happened to any of these states will be reflected to its significant other. This phenomenon is called **Quantum Entanglement**[4,11,12] Shared states are not useful to communicate.

Correlating certain sequences of statistics through entanglement can improve the quality of solution sets for a specific set of problems. In terms of data processing, linear regression techniques are completely useless for non-linear datasets and even the best Neural Network technique will poorly perform.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Suppose a N words dataset

$$V = \{w_1, w_2, \ldots, w_N\}$$

With word embedding matrix

$$\mathbf{W} \in \mathbb{R}^{N \times D}$$

Having $1, \ldots, N$ D-dimensional word vectors in each row is non-linear. Often, D=300, while N will range from hundreds to millions.

Considering a generic correlation coefficient (r), word embeddings are either be treated as N observations for a distribution with D-variant like gaussian or fitting a mixture model to cluster embeddings. Either way, semantics are the only consideration skipping the influence or relation between words. Dependency trees can define those relationships, but they are very complex at defining explicit word to word fitting. Here is where quantum shows its supremacy. Computation and processing differences are shown in table 2.

Table2. COMPUTATIONAL COMPARISON BETWEEN CLASSICAL AND QUANTUM COMPUTERS [5,11]

Classical Computer	Quantum Computer
Build with large scale multi-purpose CPU	High speed parallel computation based on quantum mechanics
Information is based on voltage/charge	Information is based on electron direction
Information processing is carried out by logical gates	Information is processed by quantum gates in parallel
Circuit behaviour is	Circuit behaviour is

based on newton's physics	governed by quantum mechanics
Operations are Boolean algebra	Operations are linear algebra and represented as unitary matrix
Descrete states (0,1)	Continuous infinite number of states
Circutes are faster and are easy to implement	Circutes are slow, fragile and use microscopic technologies

Quantum computing models are sequenced through quantum gates, unlike classical computers circuits, repeated computations on the same input will not lead to the same output. Generally, circuits are wires composed in a network carrying bit information and perform elementary operation on them as show in Fig 2.

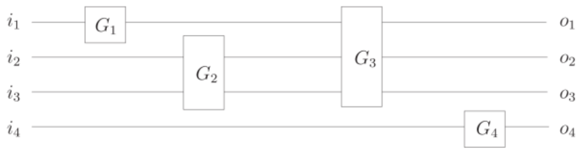


Fig 2. Circuit Diagram

A Quantum circuit system can be theorized as classical reversable system with Hadamard gate, a typical 5-qubit quantum register can be observed in Fig 3. A CCNOT gate can also be used in place of Hadamard gate[12] As data passes through qubits, we need to keep track of all qubits even they affect a part or subset of all qubits and entanglement also should be tracked as it may cause randomization.

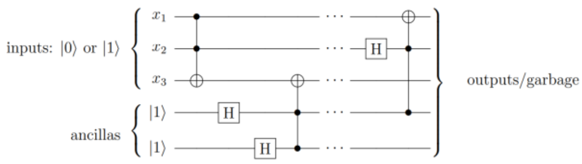


Fig 3. A 5-qubit quantum register

$$\alpha|00000\rangle + \beta|00001\rangle + \gamma|00010\rangle + \cdots + \omega|11111\rangle$$

Analysis through the circuit can be theorized as a set of rules [13],

- Hadamard Gates follows an input output behaviour $\boxed{H}$ and

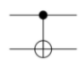
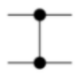
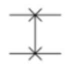
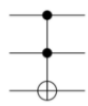
$$|0\rangle \mapsto \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle \ and \ |1\rangle \mapsto \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle$$

- CCNOT gate behaves exactly like other classical reversible gates analysed in a randomized setting
- All arithmetic operations are same except probability is defined as magnitude or amplitude
- At the output level, measurement is taken from the probabilistic rule i.e.,

$|00000\rangle$ with probability α^2
$|00001\rangle$ with probability β^2
...
$|1111\rangle$ with probability ω^2

Along with Hadmard Gate, there are exclusive quantum gates as show in table 3 and are represented as unitary matrix.

Table3. COMMONLY USED LOGIC GATES IN QUANTUM COMPUTING [12]

Operator	Gate	Unitary matrix
Hadamard	H	$\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}$
Pauli-X	X	$\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$
Pauli-Y	Y	$\begin{bmatrix}0 & -i\\i & 0\end{bmatrix}$
Pauli-Z	Z	$\begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}$
Phase	S	$\begin{bmatrix}1 & 0\\0 & i\end{bmatrix}$
CNOT, CX		$\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&0&1\\0&0&1&0\end{bmatrix}$
CZ		$\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&-1\end{bmatrix}$
SWAP		$\begin{bmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{bmatrix}$
CCNOT, CCX, TOOF		$\begin{bmatrix}1&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0\\0&0&1&0&0&0&0&0\\0&0&0&1&0&0&0&0\\0&0&0&0&1&0&0&0\\0&0&0&0&0&1&0&0\\0&0&0&0&0&0&0&1\\0&0&0&0&0&0&1&0\end{bmatrix}$

In this project, Hadamard gate is extensively used for measuring at each transformation of grammatical groups called pre-group which grammatical structures like CFS (context-free grammar) suits perfectly for quantum composition.

III. Literature Survey

Yan Yu et.al [14], have demonstrated a quantum theory based on entanglement by considering the expansion of word vectors and semantic noise formed by tensor products. The author's main intensions are to match sentences over similarities in semantics, they achieved this using the quantum's entanglement phenomenon. They have used word2vec datasets and assessed their approach on 16 custom datasets.

Prayag Tiwari and Massimo Melucci [15] have proposed a quantum binary classifier. They have used Text and Image resources to train the modal which they quote "outperformed" the state-of-art modals in categories of F-measure, Precision and Recall. They have taken the bases of SDT algorithm and upgraded it's core with quantum functions and named it Quantum Inspired Binary Classification (QIBC) which indeed out performed NB, SVM, KNN, DT.

Quantum Computing algorithms are more likely prone to error, C. Kim et.al have prosed a technique that use artificial intelligence to deal with such errors known as NISQ. They build a random noisy quantum circuit on real data with one and two qubit unitary quantum gates and inferred a neural network that calculates the quantum measure adjustment probability from the measure of unseen quantum circuit. Their proposed modal works effectively on time dependent noises with lower frequencies and also they have verified their estimations with ANN and CANN architectures.

IV. Quantum Modeling

Proposed quantum model was trained of SQuAD2.0 dataset which contains reading comprehension data based on Wikipedia articles. The previous version of this dataset contains over 10,000 questions and answers from over 500 articles. SQuAD2.0 covers these pairs with 5000 extra unanswerable questions. Many famous NLP models like FPNet, IE-Net, SA-Net-V2 and Retro-Reader modals were prepared on this SQuAD2.0 dataset[16].

Fig 4. A 3-bit multi-level Quantum Circuit

As per our main objective, relations are derived by generating S and N combinations of sentences having pairs connecting words of a sentence. From these relations, impulsive states are generated which were then given as an input to the circuit. This pre-processing step is resource extensive and is done in classical computer but critical to calculate accurate measures out of the derived grammar[17]

Considering the flow of this project, it can be divided into 3 phases[2]

1. State preparation
2. Model circuit
3. Measurement

The first step will consist of pre-processing the classical embeddings into quantum state. There are various methods in doing this, out of which amplitude encoding benefits later in deriving the native S and N conversions [18].

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, \| x \|^2 = 1 \Rightarrow |\phi(x)\rangle = \sum_{i=1}^{p} x_i|i\rangle$$

Each vector's coordinates were mapped into the amplitudes of any given quantum states. This step requires a normalized word vector but the SQuAD dataset does not support word vectoring, so a custom data pre processing step was implemented to convert a base level reading comprehension to 2-dimensional vector level, a trend that is known for extracting features our of text data [19]

The second step in this process is model circuit. This is relative to classification algorithm but in a circuit level.

$A = \{w_1, w_2, \cdots, w_i, \cdots, w_n\}$

Where w_i is the i[th] tulip word in sentence.

$|w_i\rangle = \frac{\vec{s_i}}{|\vec{s_i}|}$

Where $\vec{s_i}$ is the vector in w_i and $|w_i\rangle$ is the quantum ket in Hilbert space which is represented in a 2-dimensional vector notation.

$B = \{(w_1 w_2), (w_2 w_3), (w_3 w_4), \cdots, (w_{n-1} w_n)\}$

Adjacently paired word are entangled forming a array of tuples. Where,

$$|w_i\rangle|w_{i+1}\rangle = |w_i w_{i+1}\rangle = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_d \end{bmatrix} \otimes \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{bmatrix},$$

Thus resulting,

$$|w_i w_{i+1}\rangle = \begin{bmatrix} u_1 v_1 \\ u_2 v_1 \\ \vdots \\ u_i v_j \\ \vdots \\ u_d v_d \end{bmatrix}$$

An entangled sentence (T) is thus formed as

$$|T\rangle = \sum_{i=1}^{n-1} |w_i w_{i+1}\rangle$$

Where all the entangled coefficients simplify the sentences to one and the cosine direction of any two paired sentences represent the similarity [12,16,19].

$$\cos(|T_1\rangle, |T_2\rangle) = \frac{\langle T_1 \mid T_2 \rangle}{\| T_1\rangle| \cdot ||T_2\rangle \,|}$$

where $\langle T_1 \mid T_2 \rangle$ is inner product between $\langle T_1 \; and \; T_2 \rangle$ and $\| T_1 \rangle \,|$ is norm of T_1.

The final step Measurement will estimate the probability performing sampling measures from a distribution[2].

Fig 5. Quantum Measurements

Quantum computational measures were shown in Fig 5 contains state vector representation and Q-sphere representation. If the state measurement returns a zero indexed q-bit we can consider that as a negative impact on the sentence else, we can say there is a match between sentences.

V. Quantum Training

Modelling Procedure

1. SQuAD data pre processing
2. Dynamic Native S and N Generation
3. Constructing a Quantum Circute to process native Ss and Ns
4. Deriving Relations and Generating Vocab for each sentence
5. Combining Circuits
6. Measuring states and parameters
7. Cross Verifying the circuit
8. Training the backend instance with minimum SPSA optimizer
9. Exporting instance
10. Make predictions from trained model in classical machine

Minimum SPSA optimizer

Simulation perturbation stochastic approximation (SPSA) is a well revised algorithm for problems with multiple unknown parameters such in our case. The descent method capabilities of this algorithm find the global minima and shares this minima with other algorithms of quantum family such as simulated annealing[17,20]. As an approximation function this algorithm will only take 2 measurements out of objective function, in this case the objective function is non-linear derived from the similarity extending to its entanglement.

$$u^* = \arg \min_{\tau \in U} J(u)$$

Where $J(u)$ is the objective function and $u_{n+1} = u_n - a_n \hat{g}_n(u_n)$

Where $g(u) = \frac{\partial}{\partial u} J(u)$ evaluated at u_n converging towards 0. This further can be extended to Finate Difference Gradient Estimator (FD) or Stocastic Perturbation Gradient Estimator (SD) [21].

Training was done with mocking albert machine learning modal considering everything in lowercase of 12 per batch at 3e-5 learning rate with 11017 iterations. Maximum sequence length considered here is 384, sequences with lesser length will be padded and sequences with higher lengths will be spit and padded if necessary. On average 10 sentences took 5.71 seconds that is roughly over 30 words with combinations of 30! relations.

Result Analysis

From the training and validation measurements illustrated in fig 6, we can observe that validation starts from a very low point which is perfectly normal for any good modal as the modal only knew a part of the data. With increase in the number of epochs the validation score is kept increasing whereas the training faced downfall at epoch 3 and at epoch 5 which is again a good thing as it is learning through some padded data which are hard to comprehend towards similarities.

Starting at 0.2006, validation reached to 0.8349 which can further be improved but it may take a lot of time and may prone to overfitting as the optimizer used in this model tend to deliver a minima subject to minimum parameters. Training accuracy however started at 0.7653 reached to 0.9938 which is not a huge difference in numbers but the 0.2285 will effect the entanglement between words establishing a calculated relationship and thus we can say $\frac{\partial}{\partial u} J(u)$ is that difference 0.2285.

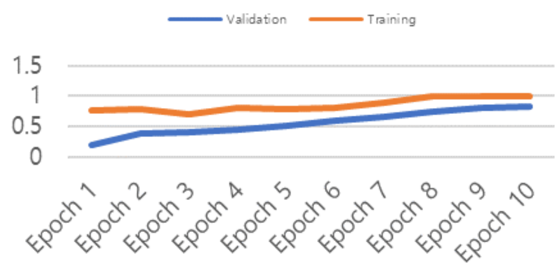


Fig 6. Training Vs Validation Graph

We can also observe that this learning process struggles till 4th epoch as the SPSA optimizer and simulated annealing may haven't formed a stable sharing bond, between them only 2 parameters are acting as an interface and it is already a known property of simulated annealing [22] that it will pass estimated values even when there is no actual data given to it. This property helps in gaining convergence faster yet struggling through the process.

Embeddings from this modal were then transferred as weights to Google's opensource ALBERT natural language processing modal to make predictions as compared with other modals it is proved to obtain optimal results with highest accuracy scores as shown in Fig 7. Any quantum modal should consider a loss while transferring or converting into a classical algorithm or modal[23]. Here in our Parameterized Contextual Modal (PCM) a loss of 8% is calculated by noisy-opt, a low noise cancellation method used in amplitude tuning.

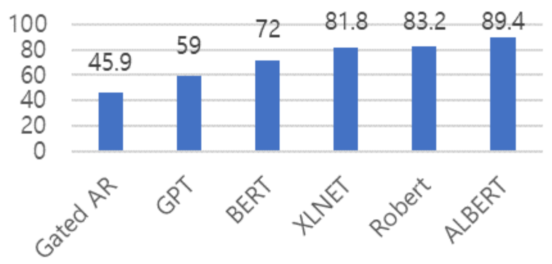


Fig 7. RACE challenge accuracy report [24]

Prediction

Question and Contexts are the bases for training, in the dataset we are using for training have a lot of related contexts for that question but when building an intelligent agent, context is hard to comprehend, thus we in this project are using Wikipedia's data to get context. For every 24hrs, 5,89,00,479 articles are being added or edited and thus resulting in different accuracies with a high precision especially for classical algorithm. The quantum modal quickly adapts to the environment thus learning through the prediction is native[10]

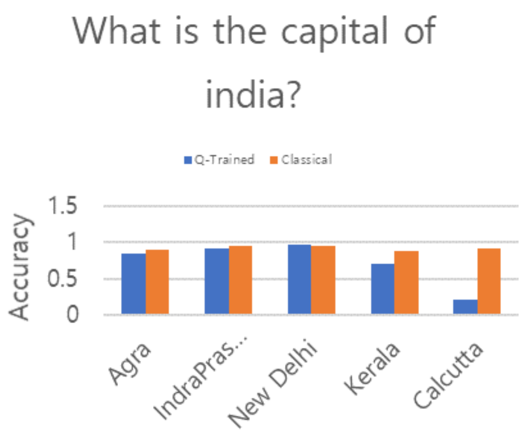


Fig 8. Results for question "What is the capital of india?"

All the answers given from the context are theoretically correct at certain point, but their scores / accuracy levels indicate the exact answer at this point of time. From the Fig 8 and Table 4, we can observe that 'New Delhi' is having score of 0.9665 and 0.94837 by Q-Trained modal and Classical modal respectively but observing closely we can see that for the rest of the answers classical scores are higher than Q-Trained. From the observation, according to classical algorithm, Indraprastha is the correct answer with accuracy of 95%.

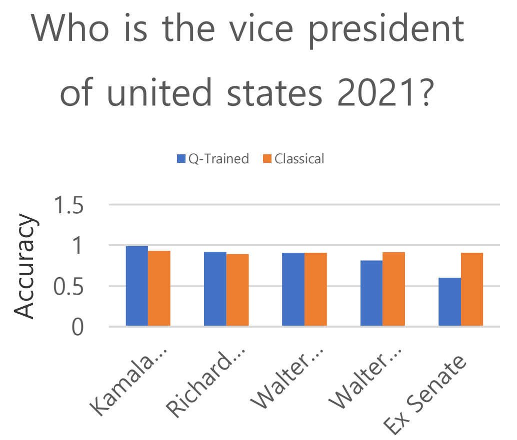


Fig 9. Result for question "Who is the vice president of united states 2021?"

In all these cases classical algorithm is exhibiting accuracy greater than 80% with a wrong answer but the intelligence of Quantum supremacy tells which is more likely to be the answer and which is less likely to be the answer. In case of another question "Who is the vice president of united states 2021?", which is specific question with details in it. From the observations of Fig 9 and Table 5. 'Kamala Harris' is the correct answer with accuracies of 99% and 93% which are highest reported accuracies for both the algorithms for wrong answers we can observe classical algorithm is showing higher accuracies than Q-Trained.

Table 4. RESULTS FOR QUESTION "WHAT IS THE CAPITAL OF INDIA?"

Answer	Q-trained	Classical
Agra	0.8428	0.8974
Indraprastha	0.9136	**0.9538**
New Delhi	**0.9665**	0.9483
Kerala	0.7071	0.8911
Calcutta	0.2089	0.9115

Table 5. RESULTS FOR QUESTION "WHO IS THE VICE PRESIDENT OF UNITED STATES 2021?"

Answer	Q-Trained	Classical
Kamala Harris	**0.9901**	**0.9311**
Richard Nixon	0.9205	0.8924
Walter Mondale	0.9070	0.9055
Walter Frederick	0.8119	0.9142
EX senate	0.5985	0.9077

VI. CONCLUSION

From the results and observations, Q-Trained Algorithm outperformed the best-in-class open-source NLP algorithm. The results from this experiment indicates a clear advantage of this approach in text analysis based on quantum computation considering circuit binding and dimensionality reduction can draw semantic information effectively without any complex computations.

In this paper, we have also showcased the quantum supremacy and usage of quantum circuit to solve problems beyond the capacity of classical computers. Language / Text processing is only a drop in the deep ocean of quantum applications. We have also presented the usage of quantum circuits to enhance or to speed up the classical algorithms in an alarming rate.

Moving forward, we attempt to implement a framework that process text on the go and will also learn simultaneously, we call that Hybrid Quantum Learning (HQL) and will also study the semantics over the weightage of the words in a sentence.

ACKNOWLEDGMENT

The Authors would like to thank IBM and IBM's Quantum Composer for this work. Views in this paper are not a reflection of policies or the team of IBM.

REFERENCES

[1] Noson S. Yanofsky. (2007). An Introduction to Quantum Computing.

[2] F. Ablayev, M. Ablayev, J. Z. Huang, K. Khadiev, N. Salikhova and D. Wu. (2020). On quantum methods for machine learning problems part II: Quantum classification algorithms. *Big Data Mining and Analytics*, 3(1), 56-67. DOI: 10.26599/BDMA.2019.9020018

[3] Pablo Arrighi, & Gilles Dowek. (2011). The physical Church-Turing thesis and the principles of quantum theory.

[4] F. Ablayev, M. Ablayev, J. Z. Huang, K. Khadiev, N. Salikhova and D. Wu. (2020). On quantum methods for machine learning problems part I: Quantum tools. *Big Data Mining and Analytics*, 3(1), 41-55. DOI: 10.26599/BDMA.2019.9020016.

[5] T. M. Khan and A. Robles-Kelly. (2020). Machine Learning: Quantum vs Classical. *IEEE Access*, 8, 219275-219294. DOI: 10.1109/ACCESS.2020.3041719.

[6] P. Qiu, X. Chen and Y. Shi. (2019). Solving Quantum Channel Discrimination Problem With Quantum Networks and Quantum Neural Networks. *IEEE Access*, 7, 50214-50222. DOI: 10.1109/ACCESS.2019.2910840.

[7] Deutsch, D. (1985). Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 400(1818), 97-117. Retrieved May 3, 2021, from http://www.jstor.org/stable/2397601

[8] S. Y. -C. Chen, C. -H. H. Yang, J. Qi, P. -Y. Chen, X. Ma and H. -S. Goan. (2020). Variational Quantum Circuits for Deep Reinforcement Learning. *IEEE Access*, 8, 141007-141024. DOI: 10.1109/ACCESS.2020.3010470.

[9] C. Kim, K. D. Park and J. -K. Rhee.(2020). Quantum Error Mitigation With Artificial Neural Network. *IEEE Access*, 8, 188853-188860. doi: 10.1109/ACCESS.2020.3031607

[10] Zeng, W., & Coecke, B. (2016). Quantum Algorithms for Compositional Natural Language Processing. *Electronic Proceedings in Theoretical Computer Science*, 221, 67–75.

[11] Ciliberto, C., Herbster, M., Ialongo, A., Pontil, M., Rocchetto, A., Severini, S., & Wossnig, L. (2018). Quantum machine learning: a classical perspective. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2209), 20170551.

[12] Abhijith J., Adetokunbo Adedoyin, John Ambrosiano, Petr Anisimov, Andreas Bärtschi, William Casper, Gopinath Chennupati, Carleton Coffrin, Hristo Djidjev, David Gunter, Satish Karra, Nathan Lemons, Shizeng Lin, Alexander Malyzhenkov, David Mascarenas, Susan Mniszewski, Balu Nadiga, Daniel O'Malley, Diane Oyen, Scott Pakin, Lakshman Prasad, Randy Roberts, Phillip Romero, Nandakishore Santhi, Nikolai Sinitsyn, Pieter J. Swart, James G. Wendelberger, Boram Yoon, Richard Zamora, Wei Zhu, Stephan Eidenbenz, Patrick J. Coles, Marc Vuffray, & Andrey Y. Lokhov. (2020). Quantum Algorithm Implementations for Beginners.

[13] YRyan O'Donnell. (2015). Lecture 1: Introduction to the Quantum Circuit Model. *Quantum Computation*, https://www.cs.cmu.edu/~odonnell/quantum15/lecture01.pdf , CMU 18-859BB.

[14] Y. Yu, D. Qiu and R. Yan. (2020). A Quantum Entanglement-Based Approach for Computing Sentence Similarity. *IEEE Access*, 8, 174265-174278. doi: 10.1109/ACCESS.2020.3025958.

[15] P. Tiwari and M. Melucci. (2019). Towards a Quantum-Inspired Binary Classifier. *IEEE Access*, 7, 42354-42372. DOI: 10.1109/ACCESS.2019.2904624.

[16] M. Lu and F. Li. (2020). Survey on lie group machine learning. *Big Data Mining and Analytics*, 3(4), 235-258. DOI: 10.26599/BDMA.2020.9020011.

[17] S. Jaf and C. Calder. (2019). Deep Learning for Natural Language Parsing. *IEEE Access*, 7, 131363-131373. DOI: 10.1109/ACCESS.2019.2939687.

[18] Konstantinos Meichanetzidis, Stefano Gogioso, Giovanni De Felice, Nicolò Chiappori, Alexis Toumi, & Bob Coecke. (2020). Quantum Natural Language Processing on Near-Term Quantum Computers.

[19] Tom Young, Devamanyu Hazarika, Soujanya Poria, & Erik Cambria. (2018). Recent Trends in Deep Learning Based Natural Language Processing.

[20] F. Tacchino et al. (2021). Variational Learning for Quantum Artificial Neural Networks. *IEEE Transactions on Quantum Engineering*, 2, 1-10. Art no. 3101110. DOI: 10.1109/TQE.2021.3062494.

[21] Y. Li, M. Tian, G. Liu, C. Peng and L. Jiao. (2020). Quantum Optimization and Quantum Learning: A Survey. *IEEE Access*, 8, 23568-23593. DOI: 10.1109/ACCESS.2020.2970105

[22] Z. Hu, J. Luo, C. Zhang and W. Li. (2020). A Natural Language Process-Based Framework for Automatic Association Word Extraction. *IEEE Access*, 8, 1986-1997, doi: 10.1109/ACCESS.2019.2962154.

[23] D. Sierra-Sosa, M. Telahun and A. Elmaghraby. (2020). TensorFlow Quantum: Impacts of Quantum State Preparation on Quantum Machine Learning Performance. *IEEE Access*, 8, 215246-215255. DOI: 10.1109/ACCESS.2020.3040798.

[24] "ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations." Google AI Blog, http://ai.googleblog.com/2019/12/albert-lite-bert-for-self-supervised.html. Accessed 3 May 2021