INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Quantum K-means Clustering and Classical k Means Clustering For Chest Pain Classification Using Qiskit

Arvind Balakumar

4 year ece-bm, *srmist,* chennai india

Abstract: Quantum computing is an upcoming field of technology which has a broad scope of increasing the current technology in a robust manner. The application area Quantum computing is very huge which ranges from battery research, protein structure research to advance computing and security areas like cryptography, quantum internet, quantum machine learning and quantum cyber security. The quantum machine learning area seems to be the most interesting because of the computing capability of a real time quantum computer. With the quantum machine learning algorithm, classical algorithm, processing speed of current quantum computers and available data in modern world we could gain insights and find patterns which were previously not possible at all.

I. INTRODUCTION

According the Moore's law its predicted that the number of transistor will double roughly every two years. Recently its being predicted that the moore's law will come to an end in this decade with rise of cost being exponentially high and the gate size of transistor being more small electrons have chance the escape from the transistor. While quantum computing does not follow the properties of moores law qubits which is a counterpart to classical bits can handle more information and in future can increase the computation exponentially following the moores law. Quantum computing increases the processing power exponentially.

The main principle of quantum computing is quantum mechanics . Quantum mechanics is the physical property applied to the scale of atoms and sub atoms .With the growth of quantum mechanics its found that the basic physical laws which govern our world macro world does not apply when we are in the quantum world or in the size of atoms and sub atoms.The main quantum computing principles used for quantum computing are Entanglement, Interference ,Superposition and measurement.

Recent research shows the practical application of entanglement with quantum teleportation the main advantage of entanglement is that the communication and transformation of information is faster than light theoretically. Superposition is a property in which the electrons or quantum particles are in at two states at one like example of classical bit the state can be 0 and 1 together at superposition. Interference is a property in which the superpostions add together and increase the amplitude and sometimes can cancel them off. Measurment is a basic property which is used to measure the state in which quantum state is present.

Qubits are analogous to the classical bits When in superposition qubits can take any state between 0 and 1 in the bloch sphere which is used for representation. Qubit is very important for quantum algorithm ,quantum machine learning algorithm. There are many machine learning algorithm like quantum support vector ,quantum kernel learning, quantum linear regression and variational eigen solver. Qubit is the basic building block for building quantum gates which is important to build quantum circuit which intern is important to build quantum algorithm and all other application in the quantum computing block. In the paper we will use quantum machine algorithm to classify data we will be using quantum k means algorithm to classify the chest pain from a heart disease dataset, and also implement the same classification using classical k means clustering algorithm.

We will be using ibm software package qiskit(quantum information software kit for quantum computation) for quantum computation. Qiskit is an open source python library which can be installed and implemented using python. We will also be using0020ibm QASM simulator to simulate a real time quantum computer which can simulate results similar to a real time quantum computer. Since current real time quantum computers have high waiting times and also increased noise ratio we will use a simulator to simulate and save the results. We will be using heart disease dataset from kaggle which is public domain dataset free for public use .We will classify the dataset into asymptotic ,non-angina pain and typical angina pain.k means algorithm is a non supervised learning algorithm. The main difference between a classical k means algorithm and quantum k means algorithm is the distance of cenroids from clusters and the calculation for centroid measurement ,the angels for measurement for the centroid in the qubit in bloch sphere for quantum k means clustering is very large and we classify the data is the large space available in the block sphere .We can classify the clusters in better distance in quantum k means clustering than classical k means clustering.

II. METHODOLOGY

We will use python and import the qiskit packages for the calculation and implementation of the program and quantum calculation. We will load the heart disease dataset as the first step and we will perform pre processing of the dataset and perform feature extraction we use the features between ages and type of chest pain for classification of the dataset. Then we will use rotational gates and find the quantum distance to find the centroid distance to classify into clusters as type of chest pain.

A. Setting up the environment

We will be using python 3.8 mostly any version above 3.7 is recommended we will install the qiskit package for basic quantum operation like creating quantum register and measuring the states and more operation.

We will also load the qiskit[machine-learning] package to perform the quantum k means algorithm. We also import numpy pandas for operation on the dataframes. We will also import sklearn to perform classical k means classification. We will also link our python environment which is set up now to ibm quantum experience to use the ibm QASM simulator.

B. Data set loading and preprocessing

We load the dataset available in public the "heart disease dataset", the dataset is available in csv format. The dataset contain 14 features like age, sex, chest pain type, thalamessia, slope of st wave, segment, ecg resting potential ,serum cholesterol, fast blood sugar, resting blood pressure and finally the target. The dataset has 303 rows or 303 records and 14 features we will be using the age and chest pain features for our classification problem. The age feature contains the age of the 303 record. The chest pain is represented by "cp" with numerical values 0 which indicate typical angina ,1 indicating non angina pain and ,2 indicating typical angina.We load the features in the quantum rotational gates using the quantum register and U3 gate. We build the Quantum circuit using the U3 gate for distance calculation

C. Visualizing the circuit

We can use the draw function in qiskit library to visualize the circuit. We have created a quantum circuit with U3 gates. The U3 gates takes a minimum of 3 input. This gate is used to perform the angle and distance calculation while computing the quantum machine learning centroid distance.
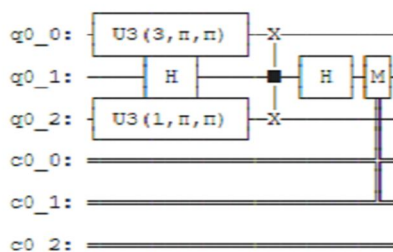

Figure 1 Quantum circuit

D. Visualizing the Superpositon State of Quantum Circuit

Since we are using the harmard gate also in our circuit the quantum circuit will be in a superposition state and the qubits value will be in superposition state.
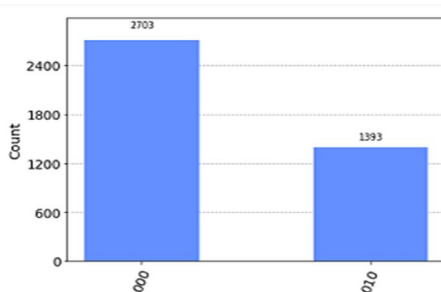

Figure 2 superposition of qubits

We could also visualize the state in which the three qubits are in present condition. We can also use an hinton plot which to find the superposition of the qubits.
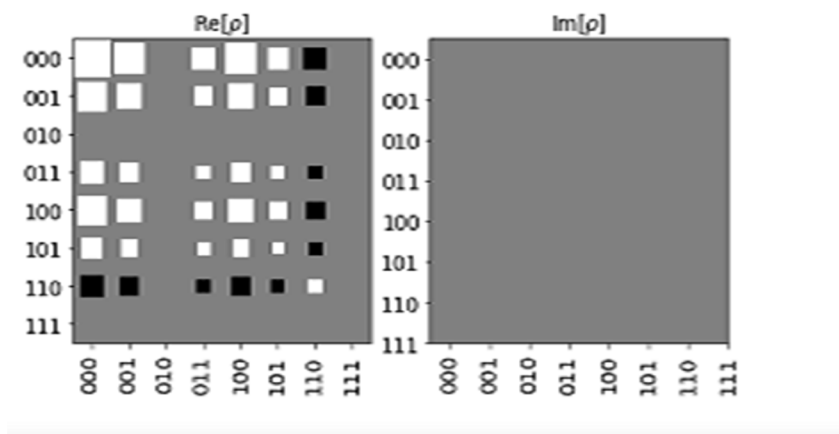

Figure 3 hinton plot

E. Classical k Means Clustering

Classical k means algorithm works on the vector quantization method .The classical k means algorithm does not allow arbitrary proximity measure. Due to the above reasons classical k means clustering uses the Euclidean distance for its centroid measurement.

The centroid are also in Euclidean space in classical computing and follows the Euclidean geometry for its centroid and its points cluster distance calculation. The main idea of k means algorithm is to minimize the square error . Updating the Euclidean distance and finding the nearest centroid and adding the data point to the cluster. The formula for finding the Euclidean distance is given below:

$$d(p,q)=\sqrt{\sum(q_i-p_i)^2}$$

Where p and q are points present in the Euclidean space and q_i and p_i are the vectors starting from the original point they are the Euclidean vectors.

F. Quantum k means Clustering

In classical k means we know that distance is the parameter which is used to classify the data points into clusters by using Euclidean distance in the presence of Euclidean space and geometry.While trying k means in quantum computing its completely different as the qubits are not of Euclidean geometry and Euclidean space.Distance will not be a good parameter to classify the qubits as the qubits are sphere and have curvature space unlike classical k means clustering. There are many parameter similar to the distance in classical k means to cluster the data point in quantum k means clustering like amplitude ,phase difference, nature of qubit

But we use the inner product between two normalized qubits and probability for states 0 and 1 as a substitute for Euclidian distance measurement to classify the points to clusters.

There are three main parts in the quantum k means algorithm which is

- Distance calculation using swap test
- Cluster updation
- Centroid updation

The formula for the distance is given by:

$$dist = d^2 r^4_{max}(P(|0\rangle-1))$$

Where d is the data points and r is the maximum limit for feature in the data

III. RESULTS
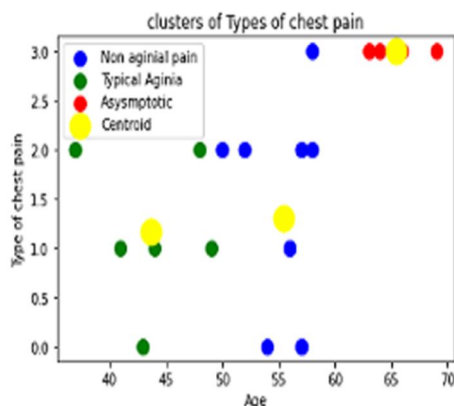
A. Classical k means Clustering Output

Figure 4 classical k means clustering

The data plots age in the x axis and cp in y axis which clusters the person based on their age classifies them into any three types of chest pain .The data points have are classified using classical k means clustering

B. Quantum K Means Clustering

The output of quantum k means clustering is similar to classical k means clustering but the centroid clusters distance are different
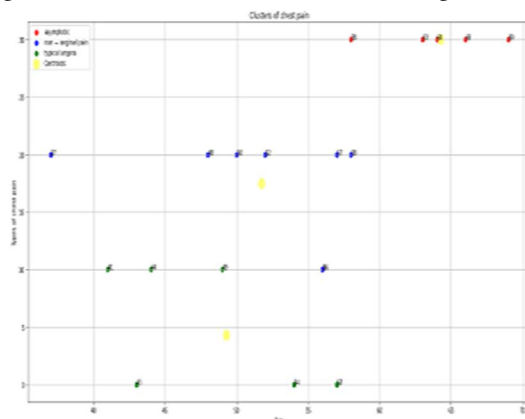

Figure 5 quantum k means clustering

IV. CONCLUSION

Quantum computers and quantum algorithm have potential high computing capability and high processing time. With the quantum machine algorithm its seen that the centroids are wide spread and have a high a high range for clusters and have all data points covered in the clusters, there are no overlapping of clusters and data points in the quantum machine learning algorithm generated clustering.

Classical k means clustering have some overlapping of clusters and datasets present in it. Though the processing time is very much similar between quantum and classical k means clustering if the simulator is replaced with real time quantum computers and the qubits are increased in upcoming years the processing time will reduce drastically for a quantum k means clustering.

REFERENCES

[1] A.J.Leggett , Quantum Mechanics: Foundations, science direct(2005), 10.1016/B0-12-369401-9/00616-1
[2] Pal, Smriti & Koley, P.K.. (2014). Quantum K-means clustering. Neural, Parallel and Scientific Computations. 22. 127-148.
[3] Wong, Renata & Garg, Tanya & Thombre, Ritu & Romo, Alberto & Niranjan, PN & Sen, Pinaki & Saggi, Mandeep & Bhatia, Amandeep. (2022). Quantum Machine Learning Algorithms. 10.1002/9781119813439.ch5.

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089   (24*7 Support on Whatsapp)