Prototyping a Quantum Clustering Algorithm for NISQ Computing

Abrar Ul Haque, Shehbaz Tariq, Haejoon Jung, and Hyundong Shin
Department of Electronics and Information Convergence Engineering, Kyung Hee University, Korea
Email: hshin@khu.ac.kr

***Abstract*—This research paper presents a novel approach to the classical k-means clustering problem by using quantum computing. The k-means algorithm is a well-known method for grouping similar data points together and our proposed quantum k-means algorithm (QKMA) aims to improve the speed and performance of the standard k-means algorithm by leveraging the principles of quantum mechanics and high throughput of FPGA. We have employed a unique amplitude embedding technique to encode our classical data into quantum states which are then fed to QKMA for faithful clustering of data points. Our quantum algorithm utilizes the inner product-based quantum perception of euclidean distance to segregate the clusters. Numerical simulations are presented to demonstrate the potential of the algorithm to provide significant speedup over the classical k-means algorithm for large datasets.**

I. Introduction

Clustering is a well-studied problem in the field of data analysis and machine learning with numerous applications such as image and speech recognition, market segmentation and bioinformatics [1]. The k-means algorithm is a widely used method for clustering similar data points together. Despite its popularity, the classical k-means algorithm can be computationally expensive for large datasets and high-dimensional data. In recent years, quantum computing has emerged as a promising technology for solving complex optimization problems faster than classical algorithms[2]. In this paper, we propose a quantum version of the k-means clustering algorithm, utilizing the principles of quantum mechanics to improve the performance and speed of the standard algorithm. Our goal is to demonstrate the potential of quantum computing in solving the k-means clustering problem and to discuss the potential of implementing the algorithm on existing and forthcoming quantum computing hardware[3]. Figure 1 outlines the architecture for classical k-means algorithm.

II. Method

This section outlines the mathematical formulation of the quantum k-means algorithm.

A. State Preparation

The classical data is mapped [4] onto the Bloch sphere by applying the following transformation on normalized coordinates of the data points.

$$\theta = (\overline{x} + 1)\frac{\pi}{2}$$

where $\overline{x} \leftarrow \frac{x}{max(\overline{x})}$

Algorithm 1: K-means Clustering

Input: a dataset of points $P = \{p_1, ..., p_n\}$, a number of clusters k
Output: centers $\{c_1, ..., c_k\}$ implicitly dividing P into k clusters
choose k initial centers C=$\{c_1, ..., c_k\}$
while *Stopping criterion has not been met* **do**
  ▷ assignment step:
  for $i = 1,...,N$ **do**
    find closest center $c_k \epsilon C$ to instance p_i
    assign instance p_i to C_k
  end
  ▷ update step:
  for i=$1,...,k$ **do**
    select c_i to be the center of mass of all points in C_i
  end
end

B. Quantum k-means Algorithm

Let the three-qubit quantum register $|0\rangle |x\rangle |y\rangle$ hold the encoded coordinates of data points between which the quantum euclidean distance is measured. We prepare the following entangled state by applying the Hadamard gate on ancilla, followed by an ancilla-controlled swap gate applied on $|x\rangle |y\rangle$:

$$\frac{1}{\sqrt{2}}(|0\rangle |x\rangle |y\rangle + |1\rangle |y\rangle |x\rangle)$$

We apply another Hadamard gate on ancilla to yield the state:

$$\frac{1}{2}(|0\rangle (|x\rangle |y\rangle + |y\rangle |x\rangle) + |1\rangle (|x\rangle |y\rangle - |y\rangle |x\rangle))$$

Application of the principles of expectation value yields that probability of the ancilla being in state $|1\rangle$ may be represented in the form of the inner product:

$$P(1_{anc}) = \frac{1}{2} - \frac{\langle x, y\rangle^2}{2}$$

Thus the probability of the ancilla being measured as $|1\rangle$ is directly correlated to the euclidean distance between the two points *x* and *y* on the Bloch sphere.

C. Implementation

We have implemented the above formulation as an FPGA accelerated quantum emulation. The hardware architecture of the same is shown in figure 1.

III. Simulation and Results

In order to evaluate the performance of the proposed quantum k-means algorithm, we conducted a series of numerical experiments using synthetic and real-world data sets. The

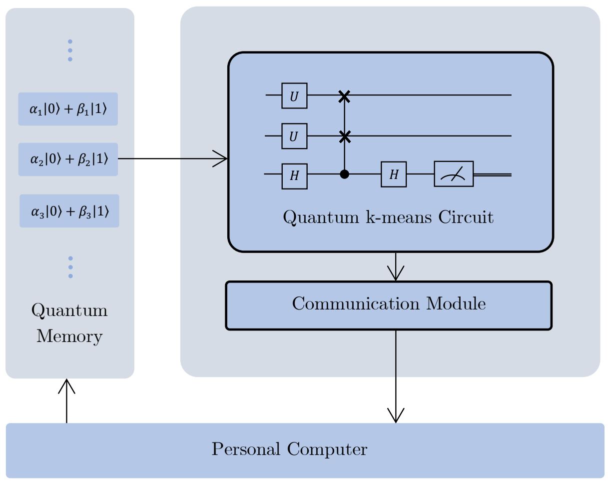

Figure 1: Hybrid implementation of quantum k-means algorithm

results were compared to those obtained using the classical k-means algorithm.

The proposed quantum k-means clustering algorithm was emulated on FPGA accelerated hybrid emulation of figure 1.

Our first set of experiments involved synthetic data sets with varying number of points and clusters. We generated 2D data sets with 800 points and 4 clusters, where the quantum k-means algorithm was able to classify the entire dataset into four clusters in 0.012 ms, while the classical k-means algorithm required an average of 5.33 ms for the same job. Figure 2 shows the clustering results for the synthetic dataset.

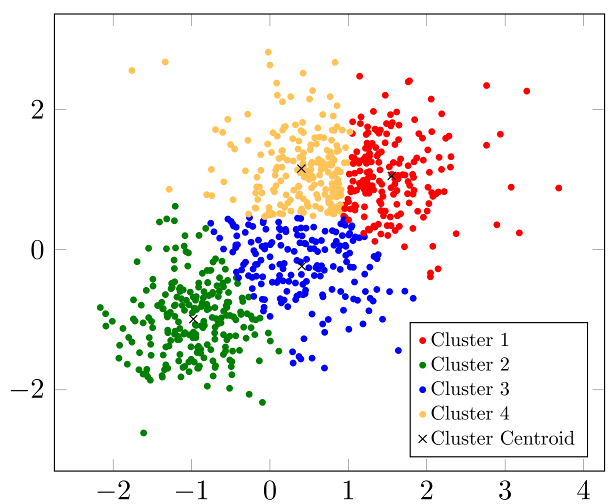

Figure 2: Clustered Synthetic Dataset

We also evaluated the performance of the quantum k-means algorithm on real-world datasets. We applied the algorithm to a dataset of air traffic passenger data and it was able to classify all the airlines with high accuracy. The results for both datasets are reported in table 1.

TABLE I
PERFORMANCE COMPARISON OF CLASSICAL K-MEANS AND ACCELERATED QUANTUM K-MEANS ALGORITHM FOR VARIOUS DATASETS

Algorithm	Implementation	Artificial	ATC
k-means	Software	5.33	45.3
Quantum k-means	Hybrid	0.012	0.109

Overall, our experiments show that the proposed quantum k-means algorithm is able to find the optimal solution for the clustering problem efficiently and accurately. The results demonstrate the potential of quantum computing to improve the performance of clustering algorithms for large and high-dimensional data sets.

IV. CONCLUSION

In this research, we presented a novel approach to the k-means clustering problem by utilizing the principles of quantum computing. We proposed a quantum version of the k-means algorithm that aims to improve the performance and speed of the standard algorithm by leveraging the high throughput of FPGA for quantum emulation. Our numerical experiments with synthetic and real-world data sets showed that our quantum k-means algorithm can significantly improve the performance of the k-means clustering problem for large and high-dimensional data sets, compared to the classical k-means algorithm.

In addition, our experiments with real-world datasets also show that the algorithm is able to produce highly accurate results. The proposed quantum k-means algorithm presents a new direction for the application of quantum computing in solving the clustering problem and opens up exciting new possibilities for future research.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1A2C2007037 and 2022R1A4A3033401) and by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-0-02046) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

REFERENCES

[1] Jing Xiao,et al., "A quantum-inspired genetic algorithm for k-means clustering," Expert Systems with Applications, vol. 37, no. 7, p. 4966-4973, 2010.

[2] Bonny, T., Haq, A. "Emulation of high-performance correlation-based quantum clustering algorithm for two-dimensional data on FPGA." Quantum Inf Process 19, 179 (2020).

[3] S. Tariq, A. Farooq, J. ur Rehman, and H. Shin, "Error analysis of gate and pulse model quantum circuits," , pp. 656–657, 2022.

[4] A. Farooq, U. Khalid, J. ur Rehman, and H. Shin, "Robust quantum state tomography method for quantum sensing," vol. 22, no. 7, p. 2669, Mar. 2022.