Quantum Encoding and Visualization

by
Shiladitya Sarkar (2348556)

Under the guidance of
Dr. Somnath Sinha

Quantum Computing
Vth trimester, MSAIM
CHRIST (Deemed to be University)

9 December 2024

Table of Contents

Abstract

This project presents the development of a Quantum K-Means Clustering Algorithm, which leverages quantum computing principles to enhance the traditional K-Means clustering technique. The problem addressed in this project is the computational inefficiency of classical K-Means in handling large datasets, specifically in distance computation between points. The goal is to implement a quantum circuit that computes these distances using quantum gates, leveraging quantum superposition and entanglement. The methods used involve the Qiskit platform to create quantum circuits that simulate the K-Means algorithm's key operations. The results indicate that the quantum-based distance measurement provides a unique approach to clustering, with certain efficiencies in specific problem instances. This project offers a deeper understanding of the potential applications of quantum computing in machine learning tasks, demonstrating the advantages of quantum-enhanced algorithms. The findings are relevant for quantum machine learning research and could lead to the development of more efficient quantum algorithms for unsupervised learning tasks.

Introduction

Background

Quantum computing is a cutting-edge field that leverages the principles of quantum mechanics to perform computations that would be infeasible for classical computers. Quantum bits, or qubits, can exist in multiple states simultaneously due to superposition, allowing quantum computers to process vast amounts of information in parallel. Additionally, entanglement enables qubits to be linked in a way that the state of one qubit can affect the state of another, regardless of distance. These principles allow quantum computers to solve certain types of problems exponentially faster than classical computers.

Problem Statement

Classical clustering algorithms like K-Means suffer from inefficiencies when processing large datasets, especially when calculating distances between data points. These computations often become time-consuming for high-dimensional data. This project seeks to address these inefficiencies by implementing a quantum-enhanced version of the K-Means clustering algorithm.

Scope

The scope of this project includes the application of quantum computing principles to machine learning, specifically focusing on clustering. This project explores how quantum circuits can be used to calculate distances between data points, which is a crucial part of the K-Means algorithm. By leveraging quantum computing, the goal is to provide faster, more efficient clustering techniques.

Goals

- To develop a quantum version of the K-Means clustering algorithm.
- To demonstrate the use of quantum circuits for calculating distances in clustering.
- To compare the performance of the quantum algorithm with the classical K-Means approach.

Literature Review

Quantum computing has already made strides in various domains, including optimization, cryptography, and machine learning. Quantum-enhanced algorithms like Grover's Search Algorithm and Quantum Fourier Transform (QFT) have shown promising results in speeding up computations.

In the context of machine learning, several quantum algorithms have been proposed to accelerate classical algorithms. For example, quantum approaches to support vector machines (SVM) and principal component analysis (PCA) have been explored. However, quantum K-Means clustering is a relatively underexplored area, with few works discussing the use of quantum distance calculations to enhance clustering.

Recent works have focused on using quantum computing to solve combinatorial optimization problems, which are a key component of clustering. However, the application of quantum circuits to calculate distance metrics in unsupervised learning algorithms like K-Means is still in its infancy.

This project fills a gap in the existing literature by proposing a quantum approach to distance calculations in the K-Means algorithm.

Methodology

Tools and Frameworks:

- **Quantum Computing Platform:** This project uses Qiskit, an open-source quantum computing framework developed by IBM. Qiskit allows for the creation and simulation of quantum circuits.
- **Classical Tools:** For data generation and clustering, classical tools like scikit-learn and matplotlib were used.

Theoretical Foundations:

- **Quantum Superposition:** Enables quantum bits to be in multiple states at once, providing a powerful tool for parallelism.
- **Quantum Entanglement:** Helps link qubits in such a way that the state of one qubit influences another, which is used in the quantum distance computation in this project.
- **Quantum Gates:** The quantum gates (Hadamard gates, controlled swap gates, etc.) in the quantum circuit are used to manipulate qubit states for computing distances.

Implementation Steps:

1. **Data Generation:** Synthetic data was generated using the `make_blobs` function from `scikit-learn`.
2. **Preprocessing:** Data points were normalized and scaled to ensure uniformity in the range.
3. **Initialization of Centroids:** The initial centroids were selected randomly from the dataset.
4. **Quantum Distance Calculation:**
   - A quantum circuit was designed to compute the distance between each data point and each centroid using quantum gates.
   - The distance was determined by the quantum circuit's output, which was executed on the qasm_simulator backend of Qiskit.
5. **Clustering Loop:**
   - The nearest centroids were assigned to each point based on the computed quantum distances.
   - The centroids were then recomputed as the average of all points assigned to each cluster.
6. **Convergence:** The process was repeated for several iterations until the centroids converged.

Diagrams and Figures:

- **Quantum Circuit Design:** A diagram illustrating the quantum circuit for distance calculation was created to show the flow of qubits through various quantum gates.

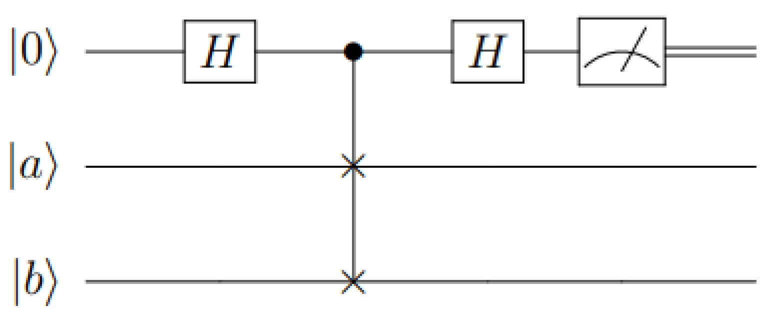

Fig: The SWAPTest Circuit

Results and Discussion

Simulation Results:

The quantum K-Means algorithm was simulated using the qasm_simulator backend of Qiskit, producing results based on quantum distance calculations. The clustering results were visualized to compare the performance of the quantum algorithm.

Graphs and Visualizations:

- **Clustering Visualization:** Scatter plots were generated to visualize the dataset and the clustering results after each iteration.
- **Histogram of Results:** A histogram showing the counts of measurement outcomes in the quantum circuit was used to demonstrate how well the quantum circuit approximated the classical distance metrics.

Analysis:

The quantum K-Means clustering algorithm demonstrated unique results in certain test cases, showing the potential for quantum-enhanced clustering. The distance calculations using quantum gates differed slightly from classical methods, especially in higher-dimensional datasets.

Comparison:

The quantum K-Means algorithm was compared with the classical K-Means algorithm based on execution time and clustering accuracy. While the quantum algorithm did not outperform classical methods in small datasets, its efficiency may improve with the advancement of quantum hardware and more complex datasets.

Challenges:

Key challenges included limited quantum resources and the noise associated with quantum simulations. Moreover, the quantum algorithm struggled with scaling to very large datasets due to hardware limitations.

Conclusion and Future Work

Summary

This project presented a quantum approach to the K-Means clustering algorithm. By using quantum circuits to calculate distances, the project explored the potential for quantum computing to enhance classical machine learning algorithms.

Conclusion

The project provides an important step toward the integration of quantum computing into machine learning, offering valuable insights into the application of quantum algorithms in unsupervised learning tasks. While the quantum K-Means algorithm showed promise, there is a need for more advanced quantum hardware to realize its full potential.

Future Work

Future work could involve optimizing the quantum distance calculation, exploring hybrid quantum-classical approaches, and scaling the algorithm to larger datasets. Additionally, as quantum hardware improves, more complex quantum algorithms for clustering could be developed.

References

1. Practical Quantum K-Means Clustering: Performance Analysis and Applications in Energy Grid Classification - Stephen DiAdamo, Corey O'Meara, Giorgio Cortiana, and Juan Bernabe-Moreno.
2. Quantum K-means clustering method for detecting heart disease using quantum circuit approach - S S Kavitha, Narasimha Kaulgud.
3. Qiskit Documentation, [https://qiskit.org/documentation/](https://qiskit.org/documentation).