Balanced K-Means Clustering on an Adiabatic Quantum Computer

Applied Quantum Machine Learning Project

Politecnico di Milano

May 16, 2021

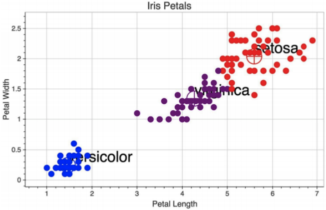
Authors:
Pierriccardo Olivieri
Francesco Piro
Matteo Sacco

Professors:
Cremonesi Paolo
Alessandro Luongo
Maurizio Ferrari Dacrema

Outline

Advantages over classical approaches

- Better targets the global solution of the training problem
- Better theoretic scalability on large datasets



Outline

- QUBO formulation and theoretical analysis
- Empirical Analysis
- Conclusions and considerations

Lloyd's algorithm

- Complexity $O(Nkdi)$ [13]
  - N number of data points
  - k number of clusters
  - d dimension of the dataset
  - i number of iterations before the algorithm converges

Scikit-learn implementation

- Complexity $O(Nkd)$ [18]

[13] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means clustering algorithm" Applied Statistics

[18] "Scikit-learn: Machine learning in python," J. Mach. Learn. Res.

Malinen et al.

- Complexity $O(N^3)$ [13]

Algorithm 1. Balanced k-means
Input: dataset X, number of clusters k
Output: partitioning of dataset.

Initialize centroid locations C^0.
$t \leftarrow 0$
repeat
  Assignment step:
    Calculate edge weights.
    Solve an Assignment problem.
  Update step:
    Calculate new centroid locations C^{t+1}
  $t \leftarrow t + 1$
until centroid locations do not change.
Output partitioning.

[21] Malinen, Mikko. (2014). Balanced K-Means for Clustering.

QUBO formulation

$$\min_{z \in \mathbb{B}^M} z^T A z$$

$$X = \{x_1, x_2, \ldots, x_N\}$$

$$\Phi = \{\phi_1, \phi_2, \ldots, \phi_k\}.$$

$$\min_{\Phi} \sum_{i=1}^{k} \sum_{x \in \phi_i} \|x - \mu_i\|^2$$

$$\min_{\Phi} \sum_{i=1}^{k} \frac{1}{2|\phi_i|} \sum_{x,y \in \phi_i} \|x - y\|^2$$

$$\min_{\Phi} \sum_{i=1}^{k} \sum_{x,y \in \phi_i} \|x - y\|^2$$

Distance matrix: D

Assignment matrix: $\hat{W}$

$$\sum_{x,y \in \phi_j} \|x - y\|^2 = \hat{w}_j^T D \hat{w}_j'$$

$$\min_{\hat{w}} \hat{w}^T (I_k \otimes D)\, \hat{w}$$

Penalties

$$\alpha \left(\hat{w}_j'^T \hat{w}_j' - N/k \right)^2$$

$$\hat{w}_j'^T \alpha F \hat{w}_j'$$

$$F = 1_N - \frac{2N}{k} I_N$$

$$\beta \left(\hat{w}_i^T \hat{w}_i - 1 \right)^2$$

$$\hat{w}_i^T \beta G \hat{w}_i$$

$$G = 1_k - 2 I_k$$

$$\min_{\hat{w}} \hat{w}^T \left(I_k \otimes (D + \alpha F) \right) \hat{w}$$

$$\hat{w}^T Q^T \left(I_N \otimes \beta G \right) Q \hat{w}$$

$$\min_{\hat{w}} \hat{w}^T \left(I_k \otimes (D + \alpha F) + Q^T \left(I_N \otimes \beta G \right) Q \right) \hat{w}$$

Hyper-parameters tuning

$$\min_{\hat{w}} \hat{w}^T \left(I_k \otimes (D + \alpha F) + Q^T (I_N \otimes \beta G) Q \right) \hat{w}$$

$$\alpha = \frac{\max(D)}{2(N/k) - 1} \qquad \beta = \max(D)$$

Converting equation into a QUBO

$$\min_{\hat{w}} \hat{w}^T \left(I_k \otimes (D + \alpha F) + Q^T (I_N \otimes \beta G) Q \right) \hat{w}$$

$$\min_W \sum_{l=1}^{k} \sum_{j=1}^{N} \sum_{i=1}^{N} \sum_{m=1}^{d} w_{il} (x_{im} - x_{jm})^2 w_{jl}$$

$$+\alpha \sum_{l=1}^{k} \sum_{j=1}^{N} \sum_{i=1}^{N} w_{il} f_{ij} w_{jl} + \beta \sum_{l=1}^{N} \sum_{j=1}^{k} \sum_{i=1}^{k} w_{li} g_{ij} w_{lj}$$

- Complexity $O(N^2 kd)$

Malinen et al.

- Complexity $O(N^3)$

Scikit-learn implementation

- Complexity $O(Nkd)$

Clustering a Benchmark Data Set I

The Iris Dataset

- Reduced due to qubit limitations on modern hardware
- Pick N/k points from $2 \leq k \leq 3$ of the data set's classes

Experiments Run

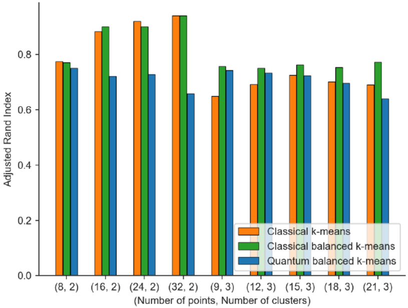
- All the 3 clustering algorithms were tested
- Experiments are run on 50 subsets of the dataset

Results

- $k = 2$
  - Trivial case, points are linearly separable
  - Classical algorithms perform better than quantum
  - Evident as the number of binary variables (Nk) increases

Clustering a Benchmark Data Set II

- $k = 3$
  - **QA** has similar performance to **Classical Balanced k-means**
  - **QA** outperforms **Scikit-Learn** implementation
  - Performance of the QA degrades as the problem size increases



Author's Conclusions

- Enhancements provided by adiabatic computers for solving **NP**-Hard or **NP**-Complete problems
- Promising result for Quantum Machine Learning
- The approach targets the global solution of the training problem **better** than the classic alternatives
- The **D-Wave 2000Q** machine
- Quantum approach partitions data with similar accuracy to the classical approaches
- The approach assumes viability as the quantum hardware improves

Future Works

- Bring the QUBO formulation to the generic k-means training problem
- Use elements of the approach to formulate quantum algorithms for similar clustering models
  - k-medoids clustering
  - fuzzy C-means clustering
- Cluster larger datasets

Can we cluster larger datasets on Advantage?

D-Wave 2000Q

- 2048 qubits
- 6,016 couplers
- 128,472 JJs

Advantage

- 5640 qubits
- 40,484 couplers
- 1,030,000 JJs

Thanks for your Attention