Quantum Collaborative K-means

Kaoutar Benlamine, Younès Bennani, Nistor Grozavu and Basarab Matei
Université Sorbonne Paris Nord, LIPN UMR CNRS 7030
99 av. J-B Clément, 93430 Villetaneuse, France
Firstname.lastname@lipn.univ-paris13.fr

***Abstract*—Recently, more researchers are interested in the domain of quantum machine learning as it can manipulate and classify large numbers of vectors in high dimensional space in reasonable time.**

In this paper, we propose a new approach called Quantum Collaborative K-means which is based on combining several clustering models based on quantum K-means. This collaboration consists of exchanging the information of each algorithm locally in order to find a common underlying structure for clustering. Comparing the classical version of collaborative clustering to our approach, we notice that we have an exponential speed up: while the classical version takes $\mathcal{O}(K \times L \times M \times N)$, the quantum version takes only $\mathcal{O}(K \times L \times \log(M \times N))$. And comparing to the quantum version of K-means, we get a better solution in terms of the criteria of validation which means in terms of clustering. The empirical evaluations validate the benefits of the proposed approach.

***Index Terms*—Quantum machine learning, Collaborative learning.**

I. Introduction

Machine learning is a branch of artificial intelligence that aims to learn from data to give the correct predictions. It is used in a wide variety of applications to solve problems like: clustering, text mining, regression, etc. There are two main types of machine learning tasks: supervised and unsupervised machine learning. In supervised machine learning, the learner is given the labeled data and the desired output. The aim is to classify new examples based on these training sets. In unsupervised machine learning, the system aims to classify the data into different groups without prior information.

Clustering is an unsupervised learning method which aims to discover the intrinsic structures of a set of objects by forming clusters of similar units. When different clusters are collaborating, we are in the collaborative clustering [12]. This latter has a purpose to reach an agreement on the partitioning of a common dataset in order to have a better overall solution.

Recent attention has been given to quantum machine learning as it can provide an exponential speed-ups for problems involving large vectors. Therefore, quantum computing has the possibility to make machine learning solutions exponentially faster than their traditional computing counterparts. That's why a lot of contributions exploring the idea of using the advantages of quantum computing in order to improve machine learning algorithms are done [16], [14], [2]. For example, Seth Lloyd and his co-workers propose the quantum version of principal component analysis [8]. Also, some works have been devoted to the development to the quantum version of a field widely used in machine learning, which are artificial neural networks [7].

In this paper, we propose a new approach called collaborative quantum K-means that aims to collaborate different K-means models in order to get a better clustering compared to QK-means and get a speed up compared to its classical version. The empirical evaluations show the benefits of this collaboration.

The rest of paper is organized as follows. Section 2 presents the classical machine learning namely K-means and collaborative K-means. Section 3 describes quantum K-means and our proposed approach quantum collaborative K-means. Section 4 is devoted to experimental results. Finally, the conclusion summarizes our contribution.

II. Classical Machine Learning

A. Classical K-means

The K-means clustering [10] is a type of unsupervised clustering, one of the most widely used clustering method early developed by Lloyd [9]. Let $X = \{x_1, x_2, \ldots, x_N\}$ be the data set, each row vector $x_n \in \mathbb{R}^M$, $1 \leq n \leq N$ is composed of M attributes (features). The K-means clustering allows to divide the set of data X into K clusters $C = \{C_1, \ldots, C_K\}$ by minimizing the following sum of squared errors:

$$R = \sum_{k=1}^{K} \sum_{n \in C_k} \|x_n - w_k\|^2 = \sum_{k=1}^{K} \sum_{n=1}^{N} g_{nk} \|x_n - w_k\|^2 , \tag{1}$$

where $w_k \in \mathbb{R}^M$ is the centroid of the data within the cluster C_k and $G \in \mathbb{R}_+^{N \times K}$ is the binary classification matrix defined by $g_{nk} = 1$, if the data $x_n \in C_k$, and 0 otherwise. Firstly, the K-means algorithm initialize randomly K centroids $w_1, w_2, ..., w_K \in \mathbb{R}^M$. Then the algorithm usually iteratively unfolds in two phases:

1) at first we go over every point and assign each to the cluster of the nearest centroid.
2) at the second phase, the centroid of each cluster is updated.

The algorithm converges when there is no further change in the assignment of instances to clusters. The idea behind K-means clustering is very natural. It just puts every new data point you ask it to classify into the group that it is closest to.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

Algorithm 1: K-means algorithm

Input: Set of vectors $x_n \in \mathbb{R}^M, n = \{1, 2, ..., N\}$, initial centroids $w_1, w_2, ..., w_K \in \mathbb{R}^M$.

Output: The set of K clusters C_k, $|C_k|$ is the number of vectors within the cluster k.

repeat

Assignment step (clustering): Assign each data to the cluster C_{k^*}, k^* is computed by:

$$k^* = \underset{k \in \{1,2,...,K\}}{\operatorname{argmin}} \|x_n - w_k\|^2$$

Update step: For all $k = \{1, 2, ..., K\}$, update the centroid w_k of each cluster C_k by:

$$w_k = \frac{1}{|C_k|} \sum_{n=1}^{N} g_{nk} x_n$$

until a stopping condition is satisfied.

B. Collaborative K-means

The aim of collaborative clustering is to make different clustering methods collaborate, in order to reach an improvement on each partitioning of some dataset. One of the first collaborative clustering algorithm was introduced in 2002 by Pedrycz [11], [12] under the name "*Collaborative Fuzzy Clustering*" (CoFC). This method was designed for the specific case of distributed data where the information cannot be shared between the different sites. This method was based on a modified version of the Fuzzy C-Means algorithm [13]. In collaborative clustering [4], the group of algorithms solve together learning problems, affecting an individual task to each learner. Interactions are recurrent between each algorithms, responsibility is collective, the action of each algorithm is geared to the performance of the group and vice versa.

The hope is that by collaborating several clustering solutions, each one with its own bias and imperfections, we will achieve a better overall solution. There are three main types of collaboration: horizontal, vertical and hybrid collaboration [4], [6]. Horizontal collaborative clustering: all datasets describe the same observations, so all the collaborative datasets have the same number of observations but a different number of variables. Vertical collaborative clustering: all datasets have the same variables. Hybrid collaborative clustering: when we use the two approaches vertical and horizontal collaborative clustering at the same time.

As K-means is not stable, so in our approach we generate L instances of K-means and we make a collaboration between these different instances in the context of quantum learning.

Suppose that we have L instances of K-means algorithm. The general collaborative clustering scheme consists of two phases:

- Phase 1: Generating a local K-means algorithm L times on the data set X. The number of clusters would be the same for all data sets. K-means identifies K cluster centers for each data set. The objective function to minimize on the first phase is:

$$R^{(l)} = \sum_{n=1}^{N} \sum_{k=1}^{K} \left(g_{nk}^{(l)}\right)^2 \|x_n^{(l)} - w_k^{(l)}\|^2, \qquad (2)$$

where l refers to the data set where the local cluster analysis is performed $l = 1, 2, \ldots, L$.

- Phase 2: After the local phase where each instance gets an initial set of cluster centroids $w_k^{(l)}, l = 1, 2, \ldots, L$. In the second phase, we perform a collaboration between the solutions of each instance algorithm. The goal is that after the collaboration, if an observation of the $X^{(l)}$-th data set is projected onto the l'-th centroid of the l-th instance algorithm, then the same observation of the $X^{(l')}$-th data set is projected on the same l' centroid. The collaborative function to minimize is:

$$C^{(l)} = \sum_{\substack{l'=1, \\ l' \neq l}}^{L} \beta_{l,l'} \sum_{n=1}^{N} \sum_{k=1}^{K} \left(g_{nk}^{(l)} - g_{nk}^{(l')}\right)^2 \|w_k^{(l)} - w_k^{(l')}\|^2, \qquad (3)$$

where $\beta_{l,l'}$ is the weight of the pairwise collaborative $C^{(l)}$ term between the instances l and l', and it's a non-negative coefficient that represents the intensity of collaboration whose value is provided by the user.

Therefore, the global objective function to minimize is given in the following form:

$$R_{Coll} = \sum_{l=1}^{L} \left(R^{(l)} + C^{(l,l')}\right). \qquad (4)$$

The objective function R_{Coll} consists of two terms. The first term represents the sum of objective functions used by L instances of K-means clustering. The second term reflects the impact of the clustering structures found in other instances. To obtain the update rule for the prototypes let us consider:

$$R_{Coll}^{(l)} = \sum_{n=1}^{N} \sum_{k=1}^{K} \left(g_{nk}^{(l)}\right)^2 \|x_n^{(l)} - w_k^{(l)}\|^2 + \sum_{\substack{l'=1, \\ l' \neq l}}^{L} \beta_{l,l'} \sum_{n=1}^{N} \sum_{k=1}^{K} \left(g_{nk}^{(l)} - g_{nk}^{(l')}\right)^2 \|w_k^{(l)} - w_k^{(l')}\|^2. \qquad (5)$$

In the calculations of the prototypes, the necessary condition for the minimum of the objective function is in the form:

$$\frac{\partial R_{Coll}^{(l)}}{\partial w_{mk}^{(l)}} = 0$$

Recall that, by definition we have:

$$\|\mathbf{x}_n^{(l)} - \mathbf{w}_k^{(l)}\|^2 = \sum_{m=1}^{M} (x_{mn}^{(l)} - w_{mk}^{(l)})^2. \qquad (6)$$

By using (6) and by differentiation with respect to $w_{mk}^{(l)}$ we get:

$$\frac{\partial R_{Coll}^{(l)}}{\partial w_{mk}^{(l)}} = -2\sum_{n=1}^{N}\left(g_{nk}^{(l)}\right)^2 (x_{mn}^{(l)} - w_{mk}^{(l)}) - 2\sum_{\substack{l'=1,\\ l'\neq l}}^{L}\sum_{n=1}^{N}\beta_{l,l'}(w_{mk}^{(l)} - w_{mk}^{(l')})\left(g_{nk}^{(l)} - g_{nk}^{(l')}\right)^2 \tag{7}$$

Therefore, the straightforward computation gives the following update rule for centroids:

$$w_k^{(l)} = \frac{\sum_{n=1}^{N}\left(g_{nk}^{(l)}\right)^2 x_n^{(l)} + \sum_{\substack{l'=1,\\ l'\neq l}}^{L}\sum_{n=1}^{N}\beta_{l,l'}\left(g_{nk}^{(l)} - g_{nk}^{(l')}\right)^2 x_n^{(l)}}{\sum_{n=1}^{N}\left(g_{nk}^{(l)}\right)^2 + \sum_{\substack{l'=1,\\ l'\neq l}}^{L}\sum_{n=1}^{N}\beta_{l,l'}\left(g_{nk}^{(l)} - g_{nk}^{(l')}\right)^2}, \tag{8}$$

where $k = 1, \ldots, K$.

Collaborative K-means clustering algorithm is presented in Algorithm 2.

Algorithm 2: Collaborative Clustering algorithm

Phase 1: Local phase

foreach $X^{(l)}$, $l = 1$ *to* L **do**

Minimize the objective function of K-means:

$$R^{(l)} = \sum_{n=1}^{N}\sum_{k=1}^{K}\left(g_{nk}^{(l)}\right)^2 \|x_n^{(l)} - w_k^{(l)}\|^2$$

end

Phase 2: Collaboration phase

foreach $X^{(l)}$, $l = 1$ *to* L **do**

Minimize the objective function of collaborative clustering:

$$R_{coll}^{(l)} = \sum_{n=1}^{N}\sum_{k=1}^{K}\left(g_{nk}^{(l)}\right)^2 \|x_n^{(l)} - w_k^{(l)}\|^2 + \sum_{\substack{l'=1,\\ l'\neq l}}^{L}\beta_{l,l'}\sum_{n=1}^{N}\sum_{k=1}^{K}\left(g_{nk}^{(l)} - g_{nk}^{(l')}\right)^2 \|w_k^{(l)} - w_k^{(l')}\|^2$$

Update the centroids by using Equation (8).

end

III. Quantum Machine Learning

A. Quantum K-means

In this section, we show the different steps of QK-means: first of all, we transform the data to quantum states. Then, we compute the distance between each state and centroids. Indeed, the estimation of the distance between states in the context of quantum learning is totally different from the classical learning.

While in the classical version the distance computation involve coordinates, in quantum learning we need to define the distance with respect to the probabilistic nature of qubits. In quantum learning notions as phase differences or amplitudes of different probabilities are easy to measure and quantify, unfortunately the estimation of the distance between two states can not be directly defined because of its instability.

More precisely, let's consider that w and x are two position vectors corresponding to some centroid cluster in the available data and to some new data point for which we want to decide the cluster assignment, respectively. In the quantum context, we associate to those two vectors two quantum states denoted by $|\phi\rangle$ and $|\psi\rangle$, we will therefore estimate the so-called distance between the states $|\phi\rangle$ and $|\psi\rangle$. As already mentioned this can not be done directly.

In a nutshell, our strategy detailed below, consist to associate the inner product of x and w to the probability that some ancillary qubit is measured on the state 0.

Then we starts by considering an ancillary qubit $|0\rangle$ and the two quantum states $|\phi\rangle$ and $|\psi\rangle$. In these two quantum states are stored the normalized position vector x and w. Since we are in the quantum context the distance between the state $|\phi\rangle$ and the state $|\psi\rangle$ entangled with the ancillary qubit $|0\rangle$ will give us the distance between x and w.

To this end, we apply a Hadamard gate to the ancillary qubit $|0\rangle$, to define the superposition. After that, we use the entanglement between the states $|\phi\rangle$ and $|\psi\rangle$ and the concerning ancillary qubit, and then we use a swap gate controlled on the ancillary qubit between the quantum state $|\phi\rangle$ and $|\psi\rangle$.

After that, we apply another Hadamard gate to the ancillary qubit, and finally we measure the ancillary qubit. In this way we recover the inner product of x and w as the ancillary probability, see Equation (9) below.

Once, we have computed the distance between each state and centroids using swap test circuit, we assign each state to the closest cluster centroid using Grover's algorithm [5]. Finally, we update the centroids of each cluster. We detail the described strategy in the next paragraphs.

1) Data preparation and states construction: Normally, there are several methods to prepare the data and construct the states. According to [2] the method of Wiebe, Kapoor and Svore [17] gives good results in terms of clustering and also stability.

In what follows, we use this method for the preparation and the construction of the states $|\psi\rangle$ and $|\phi\rangle$ for the local phase and the states $|\gamma^{(l)}\rangle$ and $|\delta^{(l')}\rangle$ for the collaborative phase.

Given $N = 2^n$ dimensional complex vectors x and w with components $x_j = |x_j|e^{-i\alpha_j}$ and $w_j = |w_j|e^{-i\beta_j}$ respectively. Assume that $\{|x_j|, \alpha_j\}$ and $\{|w_j|, \beta_j\}$ are stored as floating point numbers in quantum random access memory.

Wiebe, Kapoor and Svore [17] suggested a representation of the states that aims to write the parameters into amplitudes of the quantum states.

With the definitions of x and w, we define the quantum states:

$$|\psi\rangle = \frac{1}{\sqrt{d}} \sum_j |j\rangle \left(\sqrt{1 - \frac{|x_j|^2}{r_{max}^2}} e^{-i\alpha_j} |0\rangle + \frac{x_j}{r_{max}} |1\rangle \right) |1\rangle$$

$$|\phi\rangle = \frac{1}{\sqrt{d}} \sum_j |j\rangle |1\rangle \left(\sqrt{1 - \frac{|w_j|^2}{r_{max}^2}} e^{-i\beta_j} |0\rangle + \frac{w_j}{r_{max}} |1\rangle \right),$$

where $j = \{1, ..., n\}$, and r_{max} is an upper bound on the maximum value of any feature in the dataset. The input vectors are d-sparse, i.e., contain no more than d non-zero entries.

The idea behind this algorithm is to adjoin an ancillary qubit to the states creating an entangled state $|\psi\rangle$. The bigger difference between the states $|x_n\rangle$ and $|w_k\rangle$, the more entangled the resulting state is, and therefore we can use this entanglement to estimate the distance between vectors [3] .

2) Fidelity as a similarity measure of quantum states: Swap test circuit allows to compare two quantum states. Thus, we apply this circuit to compute the distance between quantum states and cluster centroids.

Similarity measure between two quantum states $|\psi\rangle$ and $|\phi\rangle$ could be done using the fidelity $|\langle\psi|\phi\rangle|$, see [1]. This fidelity can be obtained through the quantum swap test circuit described as follows:

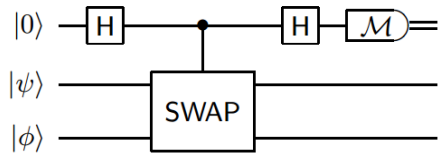

Fig. 1. Swap test circuit

The swap test circuit is composed of two Hadamard gate and a Control-Swap gate, the control qubit is on the state:

$$|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}$$

After applying the controlled swap test, we get:

$$\frac{|0\rangle |\psi\rangle |\phi\rangle + |1\rangle |\phi\rangle |\psi\rangle}{\sqrt{2}}$$

By applying the second Hadamard gate, we obtain:

$$\frac{|0\rangle (|\psi\rangle |\phi\rangle + |\phi\rangle |\psi\rangle) + |1\rangle (|\psi\rangle |\phi\rangle - |\phi\rangle |\psi\rangle)}{2}$$

While measuring the ancillary qubit, we get:

$$\begin{aligned} P(|0\rangle) &= \left| \frac{1}{2} \langle 0|0\rangle |\psi\rangle |\phi\rangle + |\phi\rangle |\psi\rangle \right|^2 \\ &= \frac{1}{4} ||\psi\rangle |\phi\rangle + |\phi\rangle |\psi\rangle|^2 \\ &= \frac{1}{2} + \frac{1}{2} |\langle\psi|\phi\rangle|^2 \end{aligned}$$

After measurement, we have $P(|0\rangle) = \frac{1}{2} + \frac{1}{2} |\langle\psi|\phi\rangle|^2$. A probability of $1/2$ shows that the two quantum states $|\psi\rangle$ and $|\phi\rangle$ are orthogonal, while a probability of 1 indicates that they have maximum overlap.

Using the swap test in our case, the distance between the states and cluster centroid is evaluated by:

$$d_q(|x\rangle, |w\rangle) = d^2 r_{max}^4 (2P(|0\rangle) - 1) \tag{9}$$

3) Cluster assignment: After computing the distance between each training state and each cluster centroid. We assign each state $|x_n\rangle$ to the closest centroid $|w_k\rangle$ by using Grover's algorithm [5]. More precisely, we should find the solution of the following minimization problem:

$$\underset{w}{\operatorname{argmin}} D(|x\rangle, |w\rangle) = \underset{C}{\operatorname{argmin}} \sum_{k=1}^{K} \sum_{|x_n\rangle \in C_k} d_q^2(|x_n\rangle, |w_k\rangle) \tag{10}$$

While the best classical algorithms for a search over unordered data requires $\mathcal{O}(N)$ time, Grover's algorithm performs the search on a quantum computer in only $\mathcal{O}(\sqrt{N})$ operations, which means a quadratic speed-up over its classical version. This speed is done thanks to the superposition of states, in other words, the search is done globally, which means a significant improvement in optimization routines.

Grover's algorithm [5] performs a search over an unordered set of 2^N items to find the unique element that satisfies some conditions. Its goal is to look in the function inputs to check if the function returns true for that input. This function can be represented as a quantum oracle and could be constructed from a large number of combined quantum gates.

4) Update the centroid: Let $X = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^{N \times M}$ be the data matrix, and $G \in \mathbb{R}_+^{N \times K}$ be the binary matrix. The classical K-means update rule for the centroids is given by $w_k = \frac{1}{|C_k|} \sum_{n=1}^{N} g_{nk} x_n$. As the columns of X^T are the vectors x_n , this can be rewritten as $w_k = G_k^T X$. So, the update of the centroid of each cluster k is given by:

$$|w_k\rangle = |G_k^T X\rangle$$

where

$$|G_k\rangle = \frac{1}{\sqrt{|C_k|}} \sum_{n=1}^{N} g_{nk} |n\rangle \quad \text{and} \quad g_{nk} = \begin{cases} 1 & \text{if } x_n \in C_k \\ 0 & \text{otherwise.} \end{cases}$$

We repeat the steps above until convergence is reached, which means until the position of cluster centroid doesn't change.

The algorithm of QK-means [2] is the following:

Algorithm 3: Quantum K-means algorithm

Input: $|X\rangle = \{|x_n\rangle \in \mathbb{C}^M, n = 1, ..., N\}$, K number of clusters C_k, initial centroids of the clusters: $|w_1\rangle, |w_2\rangle, ..., |w_K\rangle$.

Output: K clusters C_k.

repeat

Assignment step (clustering): Each data is assigned to the cluster with the nearest center using Grover's search:

$$C_{k^*} \longleftarrow \{|x_n\rangle : d_q^2(|x_n\rangle, |w_{k^*}\rangle) \leq d_q^2(|x_n\rangle, |w_k\rangle), \forall k, 1 \leq k \leq K\}$$

Update step: For all $k = \{1, 2, ..., K\}$, update the centroid $|w_k\rangle$ of each cluster C_k by:

$$|w_k\rangle \longleftarrow |G_k^T X\rangle$$

until Convergence is reached

B. Quantum Collaborative K-means

In this section, we study the collaboration between several clustering results, in particular the collaboration between several models of quantum K-means.

The general quantum collaborative clustering scheme consists of two phases:

Phase 1: Generating $l = 1, 2, \ldots, L$ clusters without collaboration, using a local QK-means algorithm on the dataset as we explained previously. QK-means minimizes the following objective function $R_Q^{(l)}$.

$$R_Q^{(l)} = d^2 r_{max}^4 \, |(G_k^2)^T\rangle \, | \, \langle \phi^{(l)} | \psi^{(l)} \rangle \, |^2 \tag{11}$$

where

$$|G_k^2\rangle = \frac{1}{\sqrt{|C_k|}} \sum_{n=1}^{N} g_{nk}^2 \, |n\rangle$$

and

$$g_{nk}^2 = \begin{cases} 1 & \text{if } x_n \in C_k \\ 0 & \text{otherwise.} \end{cases}$$

Phase 2: In the second phase, we perform a quantum collaboration between different clusters. In collaborative QK-means algorithm, when a data instance is presented from the clustering l, the optimization is done to minimize the distance between that instance and the centroids of each local QK-means iteration $l' \neq l$. Therefore the pairwise collaborative $C_Q^{(l)}$ term between the QK-means clustering l and l' is:

$$C_Q^{(l)} = d'^2 r'^4_{max} \times \sum_{\substack{l'=1, \\ l' \neq l}}^{L} \beta_{l,l'} \, | \Big((G_k^{(l)} - G_k^{(l')})^2 \Big)^T \rangle \, | \, \langle \gamma^{(l)} | \delta^{(l')} \rangle \, |^2 \tag{12}$$

Where $|\gamma^{(l)}\rangle$ and $|\delta^{(l')}\rangle$ are the two states that are associated to the vectors $w^{(l)}$ and $w^{(l')}$ .The preparation of the data and the construction of the states $|\gamma^{(l)}\rangle$ and $|\delta^{(l')}\rangle$ follows the same steps as explained previously for the states $|\psi\rangle$ and $|\phi\rangle$.

$$|\gamma^{(l)}\rangle = \frac{1}{\sqrt{d'}} \sum_c |c\rangle \left(\sqrt{1 - \frac{|w_c^{(l)}|^2}{r'^2_{max}}} e^{-i\alpha_c} |0\rangle + \frac{w_c^{(l)}}{r'_{max}} |1\rangle \right) |1\rangle$$

$$|\delta^{(l')}\rangle = \frac{1}{\sqrt{d'}} \sum_c |c\rangle |1\rangle \left(\sqrt{1 - \frac{|w_c^{(l')}|^2}{r'^2_{max}}} e^{-i\beta_c} |0\rangle + \frac{w_c^{(l')}}{r'_{max}} |1\rangle \right)$$

Where $c = \{1, ..., k\}$, and r'_{max} is an upper bound on the maximum value of any feature in the dataset. The input vectors are d'-sparse.

For simplification, let's assume in what follows that $Z = d^2 r_{max}^4$ and $Z' = d'^2 r'^4_{max}$. So, the objective function to minimize is:

$$R_{QColl}^{(l)} = Z \, |(G_k^2)^T\rangle \, | \, \langle \phi^{(l)} | \psi^{(l)} \rangle \, |^2 + Z' \sum_{\substack{l'=1, \\ l' \neq l}}^{L} \beta_{l,l'} \, | \Big((G_k^{(l)} - G_k^{(l')})^2 \Big)^T \rangle \, | \, \langle \gamma^{(l)} | \delta^{(l')} \rangle \, |^2 \tag{13}$$

The objective function presented in Equation (13) is composed of two terms, the first one corresponds to the local phase which is the sum of L instances of QK-means and the second one corresponds to the collaborative phase.

The idea of collaboration is to make the algorithms share their results with the goal of getting a better clustering results, and this can be done by adding a collaborative term to constraint the similarity between clustering elements. Where $\beta_{l,l'}$ is a non-negative matrix coefficient that represents the intensity of collaboration whose value is provided by the user.

Finally, the computation of the gradient of $R_{QColl}^{(l)}$ gives the following update rule of the centroid [15]:

$$|w_k^{(l)}\rangle = \frac{Z \, |(G_k^2)^T X\rangle + Z' \sum\limits_{\substack{l'=1, \\ l' \neq l}}^{L} \beta_{l,l'} \, | \Big((G_k^{(l)} - G_k^{(l')})^2 \Big)^T X \rangle}{Z \, |(G_k^2)^T\rangle + Z' \sum\limits_{\substack{l'=1, \\ l' \neq l}}^{L} \beta_{l,l'} \, | \Big((G_k^{(l)} - G_k^{(l')})^2 \Big)^T \rangle} \tag{14}$$

Quantum collaborative K-means clustering is presented in Algorithm 4.

Algorithm 4: Quantum Collaborative Clustering algorithm

Phase 1: Local phase
foreach $X^{(l)}$, $l = 1$ *to* L **do**
Minimize the objective function of QK-means:

$$R_Q^{(l)} = Z\,|(G_k^2)^T\rangle\,|\,\langle\phi^{(l)}|\psi^{(l)}\rangle\,|^2$$

end
Phase 2: Collaboration phase
foreach $X^{(l)}$, $l = 1$ *to* L **do**
Minimize the objective function of collaborative clustering:

$$R_{QColl}^{(l)} = Z\,|(G_k^2)^T\rangle\,|\,\langle\phi^{(l)}|\psi^{(l)}\rangle\,|^2 + Z'\sum_{\substack{l'=1,\\ l'\neq l}}^{L}\beta_{l,l'}\,|\Big((G_k^{(l)} - G_k^{(l')})^2\Big)^T\rangle\,|\,\langle\gamma^{(l)}|\delta^{(l')}\rangle\,|^2$$

Update the centroids using Equation (14).
end

IV. Empirical evaluations

A. Quantum Davies Bouldin index

As a validation criteria, we use the Quantum Davies-Bouldin (QDB) index [2]. It can be calculated as follows:

$$QDB = \frac{1}{K}\sum_{k=1}^{K}\max_{k\neq k'}\frac{\delta_n(w_k)+\delta_n(w_{k'})}{\delta(w_k,w_{k'})}, \qquad (15)$$

where

$$\delta_n(w_k) = \frac{1}{|C_k|}\sum_{i=1}^{|C_k|} d_q(|x_i\rangle_{x_i\in C_k}, |w_k\rangle)$$

and

$$\delta(w_k, w_{k'}) = d_q(|w_k\rangle, |w_{k'}\rangle).$$

And K is the number of clusters, δ_n is the average distance of all elements from the cluster C_k to their cluster center w_k, $\delta(w_k, w_{k'})$ is the distance between clusters centers w_k and $w_{k'}$. This index well evaluates the quality of unsupervised clustering because it's based on the ratio of the sum of within-clusters scatter to between-clusters separation. The main objective is to evaluate the quality of the clustering. The lower the value of QDB, the better the clustering is.

B. Datasets

- *Iris* - Iris data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant.
- *Wine* - Wine is a dataset that is related to a chemical analysis of wines grown in the same region in Italy but derived from different cultivars.
- *Wisconsin Diagnostic Breast Cancer (WDBC)* - This data has 569 instances with 32 variables (ID, diagnosis, 30 real-valued input variables). Each data observation is labeled as benign (357) or malignant (212).

C. Experimental results of the collaborative approach on different datasets

We used three different datasets to compare the quantum Davies-Bouldin (QDB) index for quantum K-means before and after collaboration. These results are represented in Table 1. For example in the Iris dataset, we can notice that QDB of the collaboration shows a decreasing behaviour compared to QK-means$_2$ which means that the quality of clustering increases. However, if we compare the collaboration with QK-means$_1$, we see that QDB increases which is normal because the collaboration isn't all the time beneficial for both collaborators it could be good for one as in our case QK-means$_2$ and doesn't improve the quality of the other collaborator QK-means$_1$.

Dataset	QK-means$_1$	QK-means$_2$	QK-means$_{coll}$
Iris	[0.35,0.57]	[0.63, 1.36]	[0.47, 0.84]
wine	[0.49, 0.56]	[0.43, 0.60]	[0.47, 0.59]
Breast Cancer	[0.48, 0.77]	[0.53, 0.97]	[0.52, 0.66]

TABLE I
QK-MEANS & QK-MEANS$_{coll}$ USING QDB INDEX

Figure 2 and 3 represent the projection of iris dataset using the principal component analysis. These figures shows the classical collaborative K-means and the quantum collaborative K-means respectively. We can notice that the algorithm of collaborative QK-means has identified the different clusters which are different from each other. Therefore, the quantum collaborative K-means gives a good classification just like it's classical version, the main difference resides in that with a quantum version the clustering can be done in a time faster than the classical version.

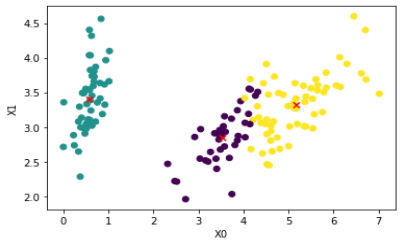

Fig. 2. Collaborative K-means clustering

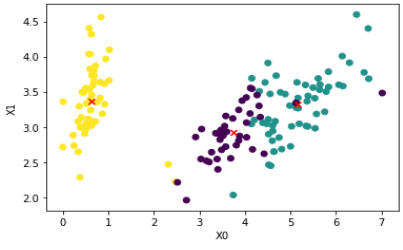

Fig. 3. Collaborative QK-means clustering

D. Computational time complexity

Let us consider the K-means problem of assigning N vectors to K clusters such that the average distance from each cluster centroid to all points of the cluster is minimized.

After randomly choosing the initial centroids, the standard method to solve this minimization problem suppose two different steps (i) assign each vector to the closest centroid and (ii) update all the centroids. This strategy is repeated until the assignment become stationary. The euclidian distance to the centroids in some N-dimensional is obtained in $\mathcal{O}(N)$. Therefore each iteration of the classical algorithm takes time $\mathcal{O}(K \times M \times N^2)$. The factor M arises since each vector is tested eventually for some reassignment. This complexity analysis is valid for classical K-means. In the case of collaboration of L instances of K-means algorithm the complexity is therefore $\mathcal{O}(K \times M \times L \times N^2)$.

Let us now analyze the quantum version of the K-means algorithm. In our strategy the assignment step in quantum version is based on the application of several quantum gates. Instead to compute the euclidean distance based on the coordinates list of two different points in $\mathbb{R}^N$, in quantum version we directly compare the two quantum states. Also finding the problem of the closest centroid is optimized since is based on Grover's algorithm [5].

The cluster assignment in the quantum context is no more a list of the different cluster assignments as in the classical K-means problem. The assignment is a quantum state which contains the different clusters labels correlated with the corresponding cluster assignments by using a quantum superposition.

Therefore the unsupervised quantum K-means has a complexity of order $\mathcal{O}(K \times \log(M \times N))$. And for the collaborative version of the quantum K-means we get at most $\mathcal{O}(K \times L \times \log(M \times N))$.

Compared to the classical version of collaborative K-means, we notice that we have an exponential speed up. Indeed, classical machine learning algorithms take time polynomial in manipulating and classifying large numbers of vectors in high dimensional spaces. On the contrary of quantum computers that can manipulate high dimensional vectors in large tensor product spaces in logarithmic time.

	collaborative K-means	collaborative QK-means
Time computational complexity	$\mathcal{O}(K \times M \times L \times N^2)$	$\mathcal{O}(K \times L \times \log(M \times N))$

TABLE II
K-MEANS COLLABORATIVE & QK-MEANS COLLABORATIVE

V. Conclusion

In the present work, our focus was on unsupervised methods for pattern clustering tasks, we started by presenting the classical K-means and collaborative K-means algorithm. Afterwards, we explained quantum K-means and our proposed approach which is quantum collaborative K-means. This approach is based on collaborating several clusters to allow the interaction between clusters. Therefore, the underlying structures and the regularities from the datasets could be detected.

The success of the quantum algorithms encourage us to propose a quantum version of the collaborative K-means algorithm that can be done in a logarithmic time in the contrary of the classical version that takes a polynomial time especially in large datasets. Our proposed approach is illustrated on various databases and the experimental results have shown very promising performance.

References

[1] Esma Aïmeur, Gilles Brassard, and Sébastien Gambs. Machine learning in a quantum world. In *Conference of the Canadian Society for Computational Studies of Intelligence*, pages 431–442. Springer, 2006.

[2] Kaoutar Benlamine, Younès Bennani, Ahmed Zaiou, Mohamed Hibti, Basarab Matei, and Nistor Grozavu. Distance estimation for quantum prototypes based clustering. In *International Conference on Neural Information Processing*, pages 561–572. Springer, 2019.

[3] X-D Cai, Dian Wu, Z-E Su, M-C Chen, X-L Wang, Li Li, N-L Liu, C-Y Lu, and J-W Pan. Entanglement-based machine learning on a quantum computer. *Physical review letters*, 114(11):110504, 2015.

[4] A. Cornuéjols, C. Wemmert, P. Gançarski, and Y. Bennani. Collaborative clustering: Why, when, what and how. *Information Fusion*, 39:81–95, 2018.

[5] Lov K Grover. A framework for fast quantum mechanical algorithms. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 53–62, 1998.

[6] N. Grozavu and Y. Bennani. Topological collaborative clustering. *Australian Journal of Intelligent Information Processing Systems*, 12(3), 2010.

[7] Sanjay Gupta and RKP Zia. Quantum neural networks. *Journal of Computer and System Sciences*, 63(3):355–383, 2001.

[8] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum principal component analysis. *Nature Physics*, 10(9):631–633, 2014.

[9] S.P. LLoyd. Least squares quantization in pcm. *Special Issue on Quantization, IEEE Tr. on Information Theory*, 28(2):129–137, 1982.

[10] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

[11] Witold Pedrycz. Collaborative fuzzy clustering. *Pattern Recognition Letters*, 23(14):1675–1686, 2002.

[12] Witold Pedrycz. *Knowledge-based clustering: from data to information granules*. John Wiley & Sons, 2005.

[13] Witold Pedrycz and Kaoru Hirota. A consensus-driven fuzzy clustering. *Pattern Recognition Letters*, 29(9):1333–1343, 2008.

[14] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd. Quantum support vector machine for big data classification. *Physical review letters*, 113(13):130503, 2014.

[15] Patrick Rebentrost, Maria Schuld, Leonard Wossnig, Francesco Petruccione, and Seth Lloyd. Quantum gradient descent and newton's method for constrained polynomial optimization. *New Journal of Physics*, 21(7):073023, 2019.

[16] Nathan Wiebe, Ashish Kapoor, and Krysta Svore. Quantum algorithms for nearest-neighbor methods for supervised and unsupervised learning. *arXiv preprint arXiv:1401.2142*, 2014.

[17] Nathan Wiebe, Ashish Kapoor, and Krysta M Svore. Quantum nearest-neighbor algorithms for machine learning. *Quantum Information and Computation*, 15, 2018.