МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2018
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2018
*Proceedings of the Forty-seventh Spring Conference*
*of the Union of Bulgarian Mathematicians*
*Borovets, April 2–6, 2018*

# ANALYSIS OF MACHINE LEARNING METHODS AND ALGORITHMS IN A QUANTUM ENTANGLEMENT-BASED ENVIRONMENT*

**Desislav A. Andreev**

Machine Learning algorithms keep finding their place in the modern software engineering and development. Although they manage to solve different tasks in the modern world, they seem to have a lack of speed. Moreover, they seem to not be capable in dealing with every possible nature of the input data. What this means is that the input must always be filtered and, if needed, normalized. This is adding complexity on actually standard and sometimes abstract tasks, where we could look at the incoming data as a vector of input values. Therefore, the quantum representation of the machine learning algorithms seems to be the next step forward. Looking even further into the technology, the research would lead to a better and closer transition to the world of the quantum mechanics. This will definitely place the question how the algorithms for machine learning will be represented if the quantum entanglement is in effect. The purpose of this paper is to look into those possibilities and propose an abstract answer of the question above by a short summary of the currently available methods.

**1. Introduction.** The main purpose of the machine learning research is to constantly analyze and provide different ways and methods not only to automate simple tasks, that are usually handled by men, but also to make the computer *think* for itself, e.g. to take an autonomous decision. Depending on the chosen topic of the machine learning there are different ways to do so – with supervision, without supervision and with reinforcement. All of them tend to find applications in real-life and they seem to manage in executing even more complex tasks – speech recognition, pattern recognition, classification, sensor management and etc. In application as these, we can easily see that the issues are handled with simple implementation of a given algorithm. Even for larger sets of data the current computing powers are sufficient.

But this is just a partial representation of the machine learning usage. When we take a look in the input with the characteristics of time series (or even time series itself), then the machine learning could be observed in a brighter sense and could be useful for finding seasoning, trends, edge-cases or even unexpected values in the incoming data. Such kind of researches are conducted, but they prove, that evaluation time is big and the

*2010 **Mathematics Subject Classification:** 37N20, 46N50, 68Q12, 81P45, 81P68, 68T05.
**Key words:** machine learning, quantum mechanics, quantum entanglement, kmeans, silhouette, clustering algorithms.

likelihood is not always satisfying [2]. This immediately leads to the reason for continuing the research of the quantum machine learning methods. Initially they step on the idea, that if the incoming data is represented in a format like in a quantum system and treated as such, this would lead to algorithms' speed-up – this gets explained in a detail below. Furthermore, the quantum entanglement is a phenomenon in the physics and an entity in the mathematics of the quantum mechanics, which cannot be ignored. Its sole purpose is to represent a system of particles, which have to be taken as a whole and not each particle as a separate one.

Such kind of an information representation is new in the quantum machine learning field and still not throughly analyzed. Although the entanglement in certain hardware setups (as in Nuclear Magnetic Resonance) is proved to be impossible already, the entanglement "boost" is still not a lost cause, because new implementations could arise [5]. This gives the opportunity for the above mentioned techniques to become even more advanced and enhancements in the further machine learning development to be applied. In order to proceed with this suggestion there must be an explanation of the translation of conventional machine learning algorithms to the quantum ones. Afterwards it must be analyzed and observed what the role of the quantum entanglement would be if it is used as an important communication resource [9]. This phenomenon is not available in the classical machine learning, so therefore this could end up being one of the best features for the quantum machine learning. In this paper we suggest a novel approach towards the quality measurement of quantum clustering by means of the *Silhouette* coefficient and proceed with a short description of what the benefit of having a quantum-entangled system would be, for the final machine learning algorithm.

**2. Methods.** Before we start with the main concepts of the quantum representation transformation actions, we need to summarize the three concepts existing in the Quantum Mechanics, that could be counted as prerequisites to the current research. In the first place this is the *quantum parallelism* [12], which unites the *superposition principle* [12] and the linearity of the quantum world in order to evaluate a single function simultaneously on arbitrarily many inputs. Immediately follows the *quantum interference* [12], that makes possible the logical paths from a given execution to affect each other in a positive way. This is actually a welcomed, so-to-say, output, because the influence between the positive outputs could affirm one another and the negative would fall out at certain point. This self-rearrangement of the system is typical expectation of the machine learning, especially when it comes to evaluations of algorithms over big data [2]. But there are also quantum states that are multi-particle, which cannot be described by the independent state of a single particle. Correlation between these states cannot be examined the classical way and therefore it is required to postulate the main source of the quantum information analysis and a powerful communication resource – the *quantum entanglement* [12].

In short the quantum-enhanced versions of classical machine learning algorithms include least-squares fitting, support vector machines, principal component analysis and deep learning [3]. In addition, the adiabatic quantum machine learning seems to work for some classes of optimization problems, and the stochastic models such as Bayesian decision theory or HMM find an elegant translation into the language of open quantum systems [11]. Still the main challenge is not only to transform these algorithms from classical to quantum world, but also to represent the classical input data in a correct

way and pass it to a given quantum device or method. The reason is that the quantum information theory must provide an understanding of how fundamental laws of nature impact the ability of physical agents to learn. In order to explain the quantum representation of the algorithms there must be first an explanation of the main entities and concepts in a quantum system:

A. Quantum bit (or q-bit) is the quantum representation of the classical term *bit*. The difference is that the q-bit exists in a superposition of states – a given electron could be at the same time in two different orbits of the same atom. Using the Dirac's notation we could note this as: $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ (with $\alpha$, $\beta \in \mathbb{C}$ and $|0\rangle$, $|1\rangle$ in the two-dimensional Hilbert space $H^2$), where $\alpha$ and $\beta$ are the amplitudes of classical states $|0\rangle$ and $|1\rangle$, maintaining the property of probability conservation given by $|\alpha|^2 + |\beta|^2 = 1$.
B. Measured state $|\psi\rangle$ – this means that either $|0\rangle$ or $|1\rangle$ is observed with probability $|\alpha|^2$ or $|\beta|^2$, respectively. This process is irreversible, because the system "collapses" to one of the proposed values and thus looses the previous value of $\alpha$ and $\beta$. All other operations in the quantum mechanics are reversible and are represented with the so-called *gates*.
C. Quantum gate – the main unit in the quantum logical circuit. Sometimes these gates have both classical and quantum representation as the Toffoli gate, for example. Each gate using a k-number of q-bits requires $2^k \times 2^k$ unitary matrix and the number of the input and output q-bits must be the same. The gate's function for a specific quantum states is evaluated through a multiplication of the state vector and the gate's matrix:
   1. for one q-bit: $V_0|0\rangle + V_1|1\rangle \rightarrow [V_0; V_1]$
   2. for two q-bits: $V_{00}|00\rangle + V_{01}|01\rangle + V_{10}|10\rangle + V_{11}|11\rangle \rightarrow [V_{00}; V_{01}; V_{10}; V_{11}]$, where the$|ab\rangle$ is the basis of the vector and the first q-bit is in $|\alpha\rangle$ state and the second – in $|\beta\rangle$.
D. Quantum random access memory (or QRAM) [7] is using $n$ q-bits to address any quantum superposition of $N$ memory cells. For a memory call $O(\log N)$ switches need to be thrown instead of the $N$ used in conventional (classical or quantum) RAM designs. This yields a more robust QRAM algorithm, as it requires in general entanglement among exponentially less gates, and leads to an exponential decrease in the power needed for addressing.
E. Quantum function computation [1] – a unitary function $f$ is computed by the quantum circuit $C$, as presented in Fig. 1. By definition it must be reversible, so it is not possible to go from $|x\rangle$ to $|f(x)\rangle$, but there must be a mapping – $|x, b\rangle$ to $|x, b + f(x)\rangle$, where the addition is performed in an appropriate finite group and the second input is a quantum register of sufficient size:

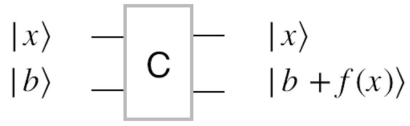


Fig. 1. An example of a quantum logical circuit

This would be required for the definition of each sub-function of a given machine learning routine. Obviously the translation process is dictated by the given classical algorithm. Our main goal would be to improve it, e.g. make it work faster, when it is transferred in the quantum world. With the usage of the the current components of a quantum circuit we should now be able to produce the logical transfer from classical to the quantum world, but only on a smaller scale, e.g. for simple logical functions and algorithms. In addition, the complex computer parts or even the computer systems themselves are not so easily transformed. The combination of quantum subsystems to bigger systems is not a simple task and cannot be based on the classical-physics representation, because of the quantum nature itself. The composite quantum system is not created only from subsystems from the same type, but they could also differ in their possible state vectors. For example, an atom is a combination of nucleons and electrons, where both of these groups of particles could be considered a separate quantum system in their own. If this is so, then we must consider that the quantum computer systems should have the ability to merge with each other in bigger ones and create more complex structures. The above mentioned quantum entanglement is exactly the "way", that describes such kind of compositions and knowing its rules we should be able to correctly scale up a given quantum computer system at least in the world of mathematics.

Having all these *instruments* we should be able to create, analyze and observe the theoretical suggestion that a larger quantum computer system could exist, but this raises another important question: Where and when does the quantum computer stop being *quantum* and become *classical*? The answer is unfortunately not that simple, because whatever is described by the quantum field theory in a given experiment, could also be described in the classical world observations, but taking into account that as the size and the complexity of the experiment (the system) increases, its state will deviate less and less from their expectation values [13]. Their predictions will approach those of the Newton mechanics due to the *Ehrenfest theorem* (1):

$$\frac{d}{dt}\langle A\rangle = \frac{1}{ih}\langle [A,H]\rangle + \left\langle \frac{\partial A}{\partial t} \right\rangle, \tag{1}$$

where $H$ is the Hamiltonian of the system, $A$ is a given quantum mechanical operator and $\langle A\rangle$ is its expectation value [12]. We will not use (1) for now.

Obviously, there could not be a clear limit, e.g. some constant number of scaling levels, which will define how much a given system should be scaled up, although the Ehrenfest equation could be re-evaluated for each case and thus could help marking the level, where the quantum system becomes pointless. In addition there has to be a clear definition of what a quantum machine learning system is, having in mind that the quantum entanglement phenomenon is happening only between two particles, but it is possible to have quantum entanglement between many q-bits, for example – the *Greenberger–Horne–Zeilinger* state [12]. The product of their entanglement will be their entangled states, e.g. the composed system's state-space, which is the tensor product of the state-spaces of at least two particles (2):

$$S_{ab} = S_a \otimes S_b. \tag{2}$$

Proceeding further, the entanglement of two q-bits would result in:

$$|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle = |\psi_1\psi_2\rangle = \alpha_1\alpha_2|00\rangle + \alpha_1\beta_2|01\rangle + \beta_1\alpha_2|10\rangle + \beta_1\beta_2|11\rangle$$

In addition to this problem, there are a few more that should be considered in the first place in order to have the quantum machine learning initiated [11]: The algorithms should be able to execute correctly the tasks for *pattern classification, pattern recognition* – to assign labels correctly to a given input set or finding a shape of patterns; *pattern completion* – adding missing values in the input dataset; *associative memory* – retrieving one of a number of stored memory vectors upon an input. Let us focus currently on the first one: If we take a look at the classification and clustering problems in the classical machine learning, it will be clear that the main concern is in finding the efficient calculations of the classical distances on a potential quantum computer. This is obviously required for the similarity measurement of two feature vectors. For the classical representation the Euclidian or the Hamming distances are used. For the quantum world there are few suggestions from [11] and [1]. All of them are basing their similarity measure on the overlap or fidelity $|\langle a|b\rangle|$ of two quantum states $|a\rangle$ and $|b\rangle$, respectively. The *swap* test is proposed, where $|a, b, 0_{\text{anc}}\rangle$ state with the two wave-functions and one ancilla register – described in [1], set to 0 is initially provided and fed to a Hadamard gate (Fig. 2):

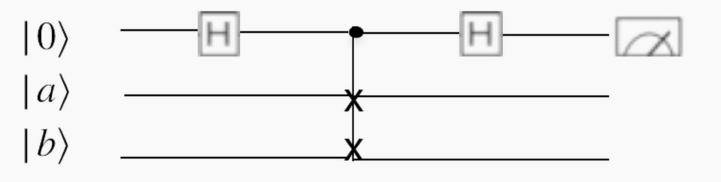


Fig. 2. *Swap* test logical circuit presentation

The following Hadamard transformation sets the ancilla into a superposition $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. The SWAP-gate on $a$ and $b$ swaps the two states if the ancilla is in $|1\rangle$. The second Hadamard gate results in state:

$$|\psi_{sw}\rangle = \frac{1}{2}|0\rangle(|a, b\rangle + |b, a\rangle) + \frac{1}{2}|1\rangle(|a, b\rangle - |b, a\rangle) \tag{3}$$

for which the probability of measuring the ground state is given by:

$$P(|0_{\text{anc}}\rangle) = \frac{1}{2} + \frac{1}{2}|\langle a|b\rangle|^2.$$

The probability of $1/2$ means that the two states do not overlap (they are orthogonal), while probability 1 indicates that they have maximum overlap.

**3. Revisiting experiments.** Proceeding further we focus our experiment on the clustering algorithms – *k-means* to be precise, because it uses the distance measure we discussed above. It is also easy to be represented mathematically both in classical and quantum machine learning. The method is NP-hard, which means it is computationally difficult, which gives further stimulation to the quantum approach – this will be discussed later. The algorithm is alternating constantly between two main steps – assign step and update step:

– **Assign** each observation $x_p$ to the cluster whose mean has the least squared Euclidean distance (nearest mean) – the observation is assigned to exactly one $S^{(t)}$:

$$S_i^{(t)} = \{x_p : \| x_p - m_i^{(t)} \|^2 \leq \| x_p - m_j^{(t)} \|^2 \; \forall j, 1 \leq j \leq k\} \tag{4}$$

– **Update** (minimize the average distance) the means to be the centroids of the observations in the new clusters:

$$m_i^{(t+1)} = \frac{1}{|s_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j. \tag{5}$$

It stops, when the assignments do not change any longer. The main idea is to partition $N$ observations into $k$ clusters. In the end the data space is partitioned in Voronoi cells. If the measurement space is $D$ dimensional the time complexity of the algorithm would be $O(D)$, where each step takes time $O(N^2D)$. Having the observations in **2.** above it would be expected, that the algorithm takes $O(N \log(ND))$. Both times $N$ is required at least once, because every vector is tested individually for the reassignment at each step [1]. Let us construct the state:

$$|\psi\rangle = \frac{1}{\sqrt{2}} \left( |\vec{x}\rangle|0\rangle + \frac{1}{\sqrt{N}} \sum_{j=1}^{N} |\vec{y}_j^S\rangle|j\rangle \right), \tag{6}$$

where $S$ is the current cluster for a set of $N$ reference vectors $\{|y_j^S\rangle\}$ of length $P$ and input vector $|\vec{x}\rangle$. The formulation of this equation results from (3) [10]. The distance can be efficiently calculated within error $\epsilon = O(\epsilon^{-1} \log NP)$. Getting back to the *swap* test and applying it over (6) we construct the following:

$$|\phi\rangle = \frac{1}{\sqrt{Z}} \left( |x||0\rangle - \frac{1}{\sqrt{N}} \sum_{j=1}^{N} |y_j^S||j\rangle \right), \text{ where } Z = |x|^2 + \left(\frac{1}{N}\right) \sum_j |y_j|^2. \tag{7}$$

The computation (7) is repeated for each cluster until a desired confidence is reached, which is usually noted by the person, working with the system. This situation could be avoided by means of an algorithm, that evaluates this confidence and marks the final iteration of the quantum *k-means*. Such method is the *Silhouette* method, which we already used in [2] and proved, that the *k-means* assigns well the received data and has a high confidence. The formula is the following:

$$s(i) = \begin{cases} 1 - a(i)/b(i), & \text{if } a(i) < b(i) \\ 0, & \text{if } a(i) = b(i) \\ b(i)/a(i) - 1, & \text{if } a(i) > b(i) \end{cases}, \tag{8}$$

where $a$ is the average distance of $i$ to the points in its cluster and $b$ – the minimal average distance of $i$ to points in another cluster, e.g. $b$ is the average dissimilarity between point $i$ and the points in the closest cluster to its cluster. From (3) and [4] we conclude that the average distance is actually the average fidelity between the points in the cluster, in our case the average fidelity between the state of the $i$-th variable in the $n$-sized cluster, which brings us to:

$$a(i) = \frac{1}{n_c} \sum_{j \in c, j=1}^{n_c} \langle F(\rho, \rho_j') \rangle_{\rho_0}, \tag{9}$$

$$b(i) = \min \frac{1}{s} \sum_{c=1}^{s} \frac{1}{n_s} \sum_{i \notin S_c, j=1}^{n_s} \langle F(\rho, \rho_j') \rangle_{\rho_0}. \tag{10}$$

Remembering that: the fidelity of two states is $\langle F \rangle = \langle F(\rho, \rho') \rangle_{\rho_0}$, where $\rho$ and $\rho'$ are

the density matrices of the $i$-th variable and any other one, respectively over all initial states $\rho_0$. Further: $s$ is the number of clusters, while $c$ is the current one, $n_S$ is the number of elements in the cluster $S$, where $i$ does not belong to them, e.g. in $S_c$ for the context of (10). From (8) the calculation of the *Silhouette* coefficient for all elements of $\vec{x}$ would be:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}, \text{ for } \forall i \in \vec{x}. \tag{11}$$

**4. Results and additional work.** As observed above in **3.** the complexity of the algorithm would be a combination of the *k-means* and the *Silhouette* routines, which takes (7) and (11) into account. We also mentioned in the explanation of (4) and (5), that the computation is difficult for big data – for high-dimensional data, which is obvious from the steps themselves. This means that the clustering algorithm will be useless with its current form, but additional studies provide workarounds [6]. The quantum approach though provides a computational complexity of $O(IN \log ND)$, where $I$ is the number of executions of the step. Because of (10) the *Silhouette* routine has at least $O(IN^2)$ complexity in the classical world, but for the QRAM execution we expect $O(N)$.

The overall evaluation complexity would be $O(I(N + N \log ND)) = O(IN \log ND)$. From the rules of asymptotic analysis we are removing the summation with $N$, but for the purposes of the current consideration we keep the multiplication by $I$. It is obvious that the evaluation is dependent on a smaller number of iterations, which is directly related to the faster discovery of the *Silhouette* coefficient closest to 1 for the given $i$. This is easier, when one takes into account the analysis in Section 2 and most importantly (2). The quantum entanglement between two q-bits will allow a simultaneous calculation of four operations – $2^n$, where $n$ is the number of q-bits. So for the algorithm of this paper the entangled environment will boost the overall evaluation to a fully logarithmic complexity, where the classical representation will be at least $O(N^2)$ as discussed above. Additional analysis will follow in the future for the relationship between the number of the entangled q-bits and the quantum algorithm complexity.

**5. Conclusion.** We presented an overview of the basic concepts in the unsupervised machine learning and their translation for quantum computers. The main properties of the quantum computation were described. A review of the basic quantum system parts was conducted and their entities' attributes were discussed. The *quantum entanglement* is being marked as the future in the scaling-up of the quantum systems, but this phenomenon is also helpful in the low-level algorithm implementations. Two classical machine learning algorithms, which take place in the *pattern recognition* problems, were translated to their quantum representation. The quantum *Silhouette* usage is a novel approach and could be used for quality measurement of the quantum clusters in runtime of the given clustering routine. It is also independent from the clustering algorithm, which makes it useful for other machine learning algorithms, when needed. The proposition of an entanglement-based environment will provide a faster computation, because of the hardware architecture itself. The performance achieved is theoretically better and there is an expectation for the existence of such quantum computer systems in the near future, so this analysis will proceed.

## REFERENCES


[1] E. Aimeur, G. Brassard, S. Gambs. Quantum Clustering Algorithms. University of Montreal, Departament of Informatics and Research, ICML '07 Proceedings of the 24th international conference on ML, pages 1–8, June 2007.

[2] D. Andreev. Pattern Recognition in Time Series. TU-Sofia, FDIBA, Master Thesis, March 2016.

[3] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, S. Lloyd. Quantum Machine Learning. Quantum Complexity Science Initiative, Institute for Quantum Computing, ICFO-The Institute of Photonic Sciences, University of Boras, Max Planck Institute of Quantum Optics, MIT, 28 November 2016.

[4] R. Cabrera, W. E. Baylis. Average Fidelity in n-Qubit system. Department of Physics, University of Windsor, Windsor, 17 January 2007.

[5] S. L. Braunstein. Separability of Very Noisy Mixed States and Implications for NMR Quantum Computing. Physical Review Letters, 1999.

[6] A. Broder, L. Garcia-Pueyo, S. Vassilvitskii, S. Venkatesan. Scalable K-Means by Ranked Retrieval. Google, 2015.

[7] V. Giovannetti, S. Lloyd, L. Maccone. Quantum random access memory. NEST-CNR-INFM & Scuola Normale Superiore, Piazza dei Cavalieri, 26 March 2008.

[8] O. Gühne, M. Kleinmann, T. Moroder. Analysing multiparticle quantum states. Universität Siegen, Naturwissenschaflich-Technische Fakultät, 24 November 2016.

[9] J. Gruska. Quantum Computing. McGraw Hill, ISBN 978-0077095031, April 2000.

[10] S. Lloyd, M. Mohseni, P. Rebentrost. Quantum algorithms for supervised and unsupervised machine learning. MIT, Research Laboratory for Electronics, Google Research, November 2013.

[11] M. Schuld, I. Sinayskiya, F. Petruccionea. An introduction to quantum machine learning. Quantum Research Group, School of Chemistry and Physics, University of KwaZulu-Natal, Durban, South Africa, National Institute for Theoretical Physics (NITheP), 11 September 2014.

[12] L. Susskind, A. Friedman. Quantum Mechanics: The theoretical minimum. Penguin Books, ISBN 978-0-141-97781-2, 2015.

[13] N. Wheeler. Remarks concerning the status and some ramifications of Ehrenfest's theorem. Reed College Physics Department, March 1998.



Desislav Andreev Andreev
Faculty for Computer Systems and Technologies
Technical University of Sofia
8, St. Kliment Ohridski Blvd
Sofia, Bulgaria
e-mail: desislav.andreev@gmail.com


## АНАЛИЗ НА АЛГОРИТМИ И МЕТОДИ ЗА МАШИННО ОБУЧЕНИЕ В СРЕДА С КВАНТОВИ СПЛИТАНИЯ


Тази статия представя кратък обзор на начините за преобразуване на класически алгоритми за машинно обучение в квантова среда и описва нововъведен метод за оценка на качеството на клъстери в квантовата среда. Анализът е част от дисертация по специалността „Системи с изкуствен интелект“ във Факултета по компютърни системи и технологии в ТУ–София.