Quantum Computing in Big Data Analytics: A Survey

Conference Paper · December 2016

DOI: 10.1109/CIT.2016.79

Tawseef Shaikh
Aligarh Muslim University

Quantum Computing in Big Data Analytics: A Survey

Tawseef Ayoub Shaikh and Rashid Ali
Department of Computer Engineering, Aligarh Muslim University, Uttar Pradesh, India
tawseef37@gmail.com, rashidaliamu@rediffmail.com

***Abstract*—Big Data is a term which denotes data that is beyond storage capacity and processing capabilities of classical computer and getting some insight from large amount of data is a very big challenge at hand. Quantum Computing comes to rescue by offering a lot of promises in information processing systems, particularly in Big Data Analytics. In this paper, we have reviewed the available literature on Big Data Analytics using Quantum Computing for Machine Learning and its current state of the art. We categorized the Quantum Machine learning in different subfields depending upon the logic of their learning followed by a review in each technique. Quantum Walks used to construct Quantum Artificial Neural Networks, which exponentially speed-up the quantum machine learning algorithm is discussed. Quantum Supervised and Unsupervised machine learning and its benefits are compared with that of Classical counterpart. The limitations of some of the existing Machine learning techniques and tools are enunciated, and the significance of Quantum computing in Big Data Analytics is incorporated. Being in its infancy as a totally new field, Quantum computing comes up with a lot of open challenges as well. The challenges, promises, future directions and techniques of the Quantum Computing in Machine Learning are also highlighted.**

Keywords—Machine Learning, Big Data Analytics, Quantum Computing, Qubits, Quantum Clustering, Quantum Artificial Intelligence

I. INTRODUCTION

Quantum computing (Quantum Information Processing) (QIP) [1] is a new computation having its roots in different interrelated disciplines. It is the application of quantum mechanics concepts in the field of information processing. QIP gains its potential power from three quantum resources that doesn't have mirror images in classical processing. Superposition principle with linearity of quantum mechanics is used in Quantum parallelism for computing a function simultaneously on arbitrarily many inputs. The logical paths of a computation to interfere in a constructive or destructive manner heading computational paths to desired results by reinforcing one another and undesired computational paths cancelling each other, is made possible by Quantum interference. Multi particle quantum states play a great role that cannot be described by an independent state for each particle.

QIP yields dramatic changes to many aspects in information technology. Since standard computing operates on bits, which are expressed in hardware as voltage levels defining the 0 or 1 values fundamental to a computing system. Nowadays with the emergence of the concepts of Quantum bits and Quantum Computation, these same quantum mechanical laws are applied to the principles of information processing. Quantum computing and quantum information sciences now define the recent trend in the information processing that consider how quantum mechanical rules impact computation and information theory. For example, quantum bits (*qubits*) are significantly different than classical bits: a qubit is defined as any of the linear superposition states $\alpha|0> + \beta|1>$, with |0> and |1> the computational basis, $\alpha$and $\beta$are amplitudes of classical states |0> and |1> such that $|\alpha|^2 + |\beta|^2 = 1$. |.> indicates a state vector for describing a quantum object. Phase $\emptyset$of a qubit is a component of the amplitude $\alpha = \acute{\alpha} e^{i\emptyset}$. Quantum theory gains its power from the fact that the squared amplitudes $|\alpha|^2 + |\beta|^2$ implies the probability to measure the qubit either in state |0> and |1>. Therefore, a qubit state is not characterized by whether it is in the `0' or `1' state, but by how likely it is to measure it in either of them. Computations can be performed on both states at the same time, a fact that is often referred to as quantum parallelism.

On the plane, a qubit is visualized as a unit vector. Also, measurements are non-reversible because the system state collapses to whichever value (|0> or |1>), so losing all memory of former amplitudes α and β. All other operations (including unitary), allowed by quantum mechanics are reversible. The input qubit | z> lies in the superposition state:

$$|z> = \frac{1}{\sqrt{2}} (|\,0> |1>) \tag{1}$$

Quantum gates are the main hardware components for carrying out the Quantum computation tasks which operates on a small number of qubits and are represented by unitary matrices. The quantum NOT gate and Hadamard gate are mainly used for this purpose. A quantum NOT gate maps | 0> to |1> and |1> to |0>respectively, and can be described by the following matrix:

$$U_{NOT} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{2}$$

When a single qubit with state $|\Psi> = \alpha\,|\,0> + \beta\,|1>$ is operated by a quantum NOT gate, it produces an output like$|\,\Psi> = \alpha\,|1> + \beta\,|\,0>$.The Hadamard gate is widespread used among the quantum gates and can be represented as:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{3}$$

The Hadamard gate transforms a qubit in the state | 0>into two states, i.e.

$$H\,|\,0> = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}}\, 0> + \frac{1}{\sqrt{2}}\, |1> \tag{4}$$

Theoretically, qubit is stored in the quantum register. A quantum register $|\chi>$, consisting of n qubits, lives in a 2^n dimensional Hilbert space. Complex amplitudes$\alpha 0,\ \alpha 1\ .\ .\ .\ .$

978-1-5090-4314-9/16 $25.00 © 2016 IEEE
DOI 10.1109/CIT.2016.79

$\alpha 2^{n}-1$ specify the register $|\chi\rangle=\sum_{i=0}^{2^n-1}\alpha i|i\rangle$ subject to normalization condition $\sum|\alpha i|^2=1$.The basic state |i>is used for the binary encoding of integer i. Two or more qubits which can always be decomposed in terms of unary and binary gates can be operated by Unitary operations. The entangling operations are also possible that create correlations between two states such that the resulting state cannot be factored into a product of the individual states. Several algorithms, including integer factorization, unstructured search and the simulation of quantum many body systems have been shown to be more efficient using qubits.

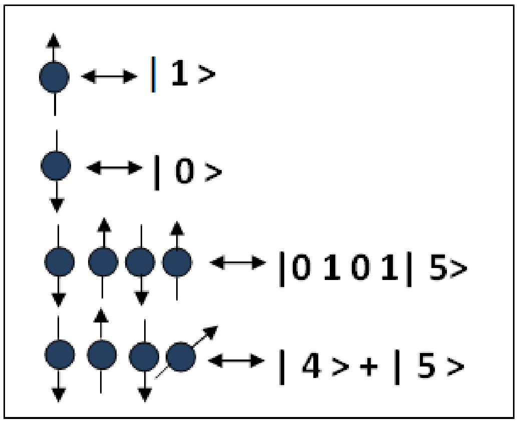


Figure 1: Qubits can be in a superposition in all the classically allowed states

So far we have not found any review/survey paper on applications of Quantum Computing in Big Data Analytics, but at the parallel side some work on application of Quantum Computing on Big Data Analytics has been done by different researchers in the same area in the recent few years. The present work may be considered as an initiative in this direction. The main contribution of the paper is to present the study of different efforts made in the area of Quantum Machine Learning for Big Data Analytics, which will become a platform for further research in the same challenging field.

The rest of this paper is organized as follows: Section II deals with the detailed description of different types of Quantum Computation techniques with their respective use in solving various types of computation problems. A brief detail of various types of the Quantum Machine Learning techniques and the application domain in which these techniques have been used is also discussed in the form of Literature review of each technique. In this section, we have categorized the Quantum machine Learning Algorithms into five different fields for better and easy understanding. Section III provides the Conclusion and the scope of the research in the given area. It also describes the potential of the Quantum Computing in Big Data Analytics and also respective challenges to harness its power in Big Data Analytics is also discussed.

II. QUANTUM COMPUTING IN MACHINE LEARNING(QUANTUM MCHINE LEARNING)

In this section, we have categorized the different Machine Learning techniques with their Quantum versions. Firstly it is Quantum Classification, Quantum Support Vector Machines, and Quantum Clustering, Quantum Reinforcement Learning followed by Quantum Searching and at the end closes with Quantum Computing for Big Data Analysis in Healthcare. The corresponding review in all subsections is carried accordingly.

Machine learning being a branch of artificial intelligence, gains its power by learning from previous experiences in order to predict about the future by making reasonable decisions, offering exceptional opportunity in the fields like computer sciences, bioinformatics, financial analysis, and robotics. The present world of Big Data creates the challenge to the machine learning with the pace of incrementally growing rate of "big data" that could become intractable for classical computers. Quantum machine learning algorithms were proposed recently that are expected to offer an exponential speedup over classical algorithms. Machine learning yields two main types of tasks [2], namely supervised and unsupervised machine learning. Supervised machine learning provides a set of training examples with features presented in the form of high-dimensional vectors and with corresponding labels for categorizing it.

Recently Lloyd, Mohseni and Rebentrost [3] that had shown that quantum computers offering good platform for manipulating vectors and matrices, could provide an exponential speed up over their classical counterparts in performing some machine learning tasks involving large vectors. For the task of assigning N-dimensional vectors to one of k clusters, each with M representative samples, a quantum computer takes time O(log(MN)).
The exponential speedup of the quantum machine learning algorithm and its potential wide applications is expected in promising applications of quantum computers [3–4], in addition to Shor's factoring algorithm [5–6], quantum simulation [7–8] and the quantum algorithm for solving linear equation systems [9].

For cluster finding and cluster assignment Lloyd et al [10], used supervised and unsupervised quantum machine learning algorithms and the work showed that quantum machine learning can provide exponential speedups over classical computers for a good number of learning tasks. A quantum computer takes time O(log(MN)) as compared with time O(poly(MN)) for the best known classical problem of assigning N-dimensional vectors.

In[11], Cai reported the experimental entanglement based classification of 2, 4, and 8 dimensional vectors to separate clusters by using a small scale photonic quantum computer which finds its use in the implementation of supervised and unsupervised machine learning in order to demonstrate the working principle of using quantum computers to classify and manipulate high dimensional vectors.

A: Quantum computing for pattern classification

Classification is a supervised learning process which identifies a set of categories (sub-populations) to which a new observation belongs to after being trained with a set of data containing observations (or instances) whose category membership is known. A new Quantum Pattern Classification algorithm for binary feature vectors similar to the distance weighted *k*- nearest neighbor method that draws on Trugenberger's proposal for measuring Hamming distance on Quantum Computer is introduced by Schuld in [12].

Schuld et al [13], give an algorithm that solves the problem of pattern classification on a quantum computer, performing linear regression effectively with least squares optimization. It runs in time logarithmic in the dimension N of the feature vectors as well as independent of the size of the training set if the inputs are given as quantum information. Instead of requiring the matrix containing the training inputs X to be sparse, it merely needs X*X to be represented by a low rank approximation.

In [14], Lu studied the quantum version of a decision tree classifier to bridge the gap between machine learning and quantum computation. Quantum entropy impurity criterion which is used to determine the node to be split is presented in the paper. A fidelity measure between two quantum states is then used and a cluster of the training data into subclasses was done so that the quantum decision tree can manipulate quantum states.

Liu et al [15], propose a new classifier having its roots in quantum computation theory. The performance test of QC was carried on two different datasets and a comparison of the performance of QC with different other classical classification methods including Support Vector Machine (SVM) and K-nearest neighbor (KNN) is carried out. The results implied that the QC outperformed both KNN and SVM on small scale raining sets, when the number of training samples is less than 50.

B: Quantum support vector machine for big feature and big data classification

Support vector machine is a powerful method for performing classification, both linear and non-linear. The classification criteria is to find the maximum margin hyper plane that divides the points with $y_j = 1$ from those with $y_j = -1$ in the case of linear support vector machines. The machine finds two parallel hyper planes having normal vector ~u, separated by the maximum possible distance 2/|~u| which separate the two classes of training data.

Rebentrost et al [16], used support vector machine for implementing an optimized linear and non-linear binary classifier on a quantum computer with exponential speedups in the size of the vectors and the number of training examples. A non-sparse matrix simulation technique to efficiently perform a principal component analysis and matrix inversion for training the data kernel matrix lies at the basic core of the algorithm.

Dynamic Quantum Clustering (DQC) is a powerful visual method working with big and high dimensionality data, is discussed by Weinstein in [17]. Its benchmark is that it exploits variations of the density of the data (in feature space) and unmasks subsets of the data that works with big, high dimensional data. A movie which showing how and why sets of data points are actually classified as members of simple clusters exhibit correlations among all the measured variables is the outcome of a DQC analysis

Rebentrost et al[18], discussed about the optimized binary classifier which can be implemented on a quantum computer, possessing logarithmic complexity in the size of the vectors and the number of training examples.

Marghny et al [19], presented the Generalized Eigen value Proximal SVM (GEPSVM)for solving the SVM complexity. Error or noise affects the data in the real world applications and working with this data is a challenging problem. In this paper an approach has been proposed to overcome this problem. This method is called DSA-GEPSVM.

Anguita et al [20], discussed the application of Quantum Computing to solve the problem of effective SVM training especially in the case of digital implementations. A comparison of the behavioral aspects of conventional and enhanced SVMs is carried out and experiments in both assynthetic and real world problems is also carried to support the theoretical analysis. The presented research at the same time differences between Quadratic-Programming and Quantum-based optimization techniques

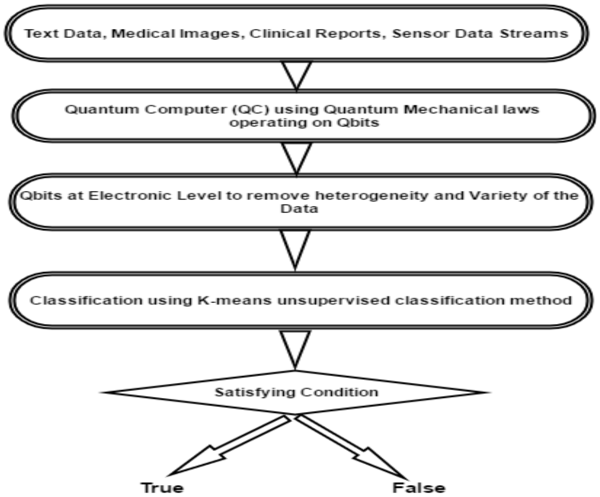
F. Quantum computing in Smart Healthcare

Smart health is the implementation of intelligent, networked technologies for constantly improving health provision for all. Among others mainly radio frequency identification (RFID), wireless sensor network (WSN), IoT (Internet of Things) and smart mobile technologies are leading this evolutionary trend. These technologies governed with the migration of the health care industry to electronic patient records and the emergence of a growing number of enabling health care technologies (e.g., wearable devices, novel biosensors and intelligent software agents, demonstrate unprecedented potential for delivering an intelligent health care in the home while at the same time reducing the cost of care. Automation of artificial intelligence into the home environment is not new. A more robust set of features, including collective intelligence algorithms, secure interactions with electronic patient records, advanced processing algorithms for physiological trend data and a host of other capabilities required for Smart health care delivery in the home. Bioinformatics research consists of voluminous, incremental and complex datasets. Kashyap et al [21], used machine learning methods in the same to handle the Variety and volume issue of Big Data.

Perez et al [22], outlines the key characteristics of big data and how medical and health informatics, sensor informatics, translational bioinformatics, and imaging informatics will get benefited from an integrated approach of integrating together different aspects of personalized information from a diverse range of data sources both structured and unstructured, covering proteomics, metabolomics, genomics as well as imaging, clinical diagnosis, and long-term continuous physiological sensing of an individual..

Gandhi et al[23], devised a neural information processing architecture inspired by quantum mechanics and incorporating the all-time known Schrodinger wave equation. The proposed architecture known as recurrent quantum neural network

(RQNN) can characterize a non-stationary stochastic signal as



time varying wave packets.

Figure 2: A framework for using Quantum Computing for Big Data Analytics in Healthcare

III. CONCLUSION

In our work, we have carried out the review of the available literature on Big Data Analytics using Quantum Computing for Machine Learning and its current state of the art. Our work categorized the Quantum Machine learning in different domains depending upon the logic of their learning. We discussed Quantum Walks, which are used to construct Quantum Artificial Neural Networks, which exponentially speed-up the quantum machine learning algorithm. We also discussed Quantum Supervised and Unsupervised machine learning and compared its benefits with respect to the Classical Supervised and Unsupervised machine learning techniques. The limitations of some of the existing Machine learning techniques and tools are also enunciated, and the significance of Quantum computing in Big Data Analytics is incorporated**.**

Quantum Machine Learning posed a hot challenge in the Information Processing since the field of Quantum Computing is still in its infancy stage because of the unavailability of Quantum Computers and necessary hardware for its `implementation, lack of proper tools and simulation environments for carrying out Quantum simulation. But a lot of progress is going on in this field and in time, it may become the treasure house for the Big data Analytics specifically in the Healthcare sector. Since Healthcare sector contains the data in lot of formats like text, image, sensor readings, and streaming data. Likewise Quantum Computing has its basic units as Quantum (Photons), so it can be worth of use to remove this heterogeneity or variety problem in the Big data, as the data in it is being analyzed at the electronic level. Once the Quantum computer hardware will be ready in the next couple of years, Quantum Computing will be the hottest topic for tackling down the Big data Analytics problems.

References

[1] SM. Barnett, "Introduction to Quantum Information", School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK, Oxford University Press, pp. 11-33.

[2] M. Mohri, A. Rostamizadeh and A. Talwalkar, "Foundations of Machine Learning", MIT Press, Cambridge, Massachusetts) 2012.

[3] S. Lloyd, M. Mohseni and P. Rebentrost, "Quantum Algorithms for Supervised and Unsupervised Machine Learning", *Quantum Physics*, Springer, Vol-1, pp. 1-11, 2013.

[4] E. Aimeur, G. Brassard and S. Gambs, "Machine Learning in an quantum world", Advances in artificial intelligence, Springer, pp. 431-442, 2013.

[5] PW. Shor, "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer (SIAM)", *Quantum Physics*, Springer,pp. 1-28, 1997.

[6] E. Lucero,R. Barends, Y. Chen, J. Kelly, M. Mariantoni, A. Megrant, PO. Malley, D. Sank, A. Vainsencher, J. Wenner, T. White, Y. Yin, AN. Cleland and JM. Martinis "Computing prime factors with a Josephson phase qubit quantum processor",Nature Physics Vol-8, pp.719-723, 2012.

[7] S. Lloyd, " Universal Quantum Simulators", *Science*, PubMed, Vol- 273, pp. 1073-1078, 1996.

[8] AA. Houck, HE. Tureci and J. Koch, "On-chip quantum simulation with superconducting circuits", *Nature Physics*, Nature, Vol- 8, pp. 292-299, 2012.

[9] AW. Harrow, A. Hassidim and S. Lloyd, " Quantum Algorithms for Solving Linear System of Equations", *Physical Review Letters*, APS physics, Vol- 103, 150502, 2009.

[10] S. Lloyd, M. Mohseni and P. Rebentrost, "Quantum algorithms for supervised and unsupervised machine learning", *Quantum Physics*, Springer ,Vol- 2, pp.1-11, 2013.

[11] D. Dong, C. Chen, H. Li and TJ Tarn, "Quantum Reinforcement Learning", IEEE Transactions on Systems Man and Cybernetics Part B: Cybernetics, Vol-38, Issue-5, pp. 1207-1220, 2008.

[12] M. Schuld, I. Sinayskiy and F. Petruccione, "Pattern Computing for Pattern Classification",Proceedings of 13th Pacific Rim International Conference on Artificial Intelligence, Gold Coast, QLD, Australia, December 1-5, 2014. Trends in Artificial Intelligence, pp.208-220, 2014, Vol- 8862, Springer International Publishing.

[13] M. Schuld, I. Sinayskiy and F. Petruccione , "Pattern classification with linear regression on a quantum computer",1601.07823, pp.1-5, 2016..

[14] S. Lu and SL. Braunstein, "Quantum decision tree classifier", Vol-13, pp.757-770, 2014, Quantum Inf Process, Springer.

[15] D. Liu, X. Yang, M. Jiang, "A Novel Text Classifier Based on Quantum Computation",Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL , 4-9 August 2013, Sofia, Bulgaria, Volume- 2, pp. 484-488.

[16] P. Rebentrost, M. Mohseni and S. Lloyd, "Quantum Support Vector Machine for Big Data Classification", Vol -14, pp. 3-8, , American Physical Society, 2014.

[17] MH. Marghny, RM. Abd El-Aziz and AI. Taloba, " Differential Search Algorithm-based Parametric Optimization of Fuzzy Generalized Eigenvalue Proximal Support Vector Machine", International Journal of Computer Applications, Vol- 108, pp.0975 – 8887, December 2014.

[18] D. Anguita, S. Ridella, F. Rivieccio and R.Zunino, "Quantum optimization for training support vector machines", Neural Networks, Elsevier ,Vol -16, pp. 763–770, 2003.

[19] E. Aımeur, G.Brassard and S. Gambs, "Quantum Clustering Algorithms", Proceedings of the 24 th International Conference on Machine Learning (ICML-07), Corvallis, pp.1-8, 2007.

[20] R. Dridi and H. Alghassi, "Homology Computation of Large Point Clouds using Quantum Annealing", Journal of Machine Learning Research. Vol-6, pp. 1-16, 2015.

[21]Heidi's Home Automation System,http://www.hometoys.com/htinews/feb99/articles/vogel/index.htm.

[22] JA. Perez, CY. Carmen, D. Robertd, TC.Stephen and GZ. Yang, "Big Data for Health", IEEE Journal of Biomedical and Health Informatics, Vol- 19, Issue- 4, , pp. 1193-1208, July 2015.

[23]M. Weinstein, F. Meirer, A. Hume, P. Sciau, G. Shaked, R. Hofstetter , E. Persi, A. Mehta, and D. Horn., "Analyzing Big Data with Dynamic Quantum Clustering", pp.1-37, U. S. DOE, Contract No. DE-AC02-76SF0051.