

Quantum assisted unsupervised data clustering on the basis of neural networks

I. D. Lazarev

*Institute of Problems of Chemical Physics of Russian Academy of Sciences,
Acad. Semenov av. 1, Chernogolovka, Moscow Region, Russia, 142432 and
Faculty of Fundamental Physical-Chemical Engineering,
Lomonosov Moscow State University, GSP-1, Moscow, Russia 119991*

Marek Narożniak

*New York University Shanghai, 1555 Century Ave, Pudong, Shanghai 200122, China and
Department of Physics, New York University, New York, NY, 10003, USA.*

Tim Byrnes

*New York University Shanghai, 1555 Century Ave, Pudong, Shanghai 200122, China
State Key Laboratory of Precision Spectroscopy, School of Physical and Material Sciences,
East China Normal University, Shanghai 200062, China
NYU-ECNU Institute of Physics at NYU Shanghai,
3663 Zhongshan Road North, Shanghai 200062, China
National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan and
Department of Physics, New York University, New York, NY 10003, USA*

A. N. Pyrkov*

*Institute of Problems of Chemical Physics of Russian Academy of Sciences,
Acad. Semenov av. 1, Chernogolovka, Moscow Region, Russia, 142432
(Dated: April 20, 2020)*

Unsupervised machine learning is one of the main directions in the area of artificial intelligence. Here, we propose a realization of quantum assisted unsupervised data clustering on the basis of artificial neural network called a self-organizing feature map. We make a proof-of-concept realization of one of the crucial parts of the approach on IBM quantum machine and show that it allows us to reduce number of calculations in a number of clusters. We compare the results with the classical algorithm on a toy example of unsupervised text clustering.

I. INTRODUCTION

The combination of big data and artificial intelligence — dubbed the fourth industrial revolution — has profoundly affected the modern economy in a plethora of different ways from robotics to agriculture [? ? ? ? ?]. Contemporary artificial intelligence methods based on neural networks also have the potential to enhance the role of novel analytical methods in science and engineering [? ? ? ? ?]. Paradoxically, the exact mechanism of why neural networks are so powerful remain unknown (in many cases it is regarded as a black box). It has been speculated that limits to the neural network approach based on the computational power of von Neumann architecture are being approached, and improvements appear only due to heuristic escalation of complexity [? ? ? ?]. In particular, the self-organizing feature map (SOFM) [? ? ? ?], is a type of artificial neural network (ANN) that is trained in an unsupervised manner. SOFMs are used in many areas [? ? ? ? ? ? ? ?] and in comparison with many other artificial neural networks they apply competitive learning and preserve the topological properties of the input space [?]. The SOFMs with a

small number of nodes are similar to the K-means algorithm but for larger SOFMs they represent data in a fundamentally topological way that allows one to do dimensionality reduction. Once it is trained, the map can classify a vector from the input space by finding the node with the smallest distance metric.

Meanwhile, there has been much interest recently in applying quantum computing techniques to machine learning [? ? ? ? ?]. In the early stage, the main focus in QML was on the use of quantum computers to perform some basic linear algebra subroutines faster (such as Fourier transform and solving systems of linear equations [? ? ?]), open up the possibilities for exponential speed-ups utilizing quantum phenomena such as quantum entanglement and quantum superposition over classical counterparts [? ?] and, in the result, quantum versions were developed for linear regression, principal component analysis, support vector machine, K-means algorithm and so on [? ? ? ? ?]. Recently, a number of shifts occurred and the hottest trend migrated to developing quantum neural networks [? ? ?]. The shift was inspired with the progress in experimental quantum computing when it became possible to play around with parametrized quantum circuits, whose parameters behave much like the weights of a neural network [?]. In particular, quantum algorithms for training and evaluat-

* Email address:pyrkov@icp.ac.ru

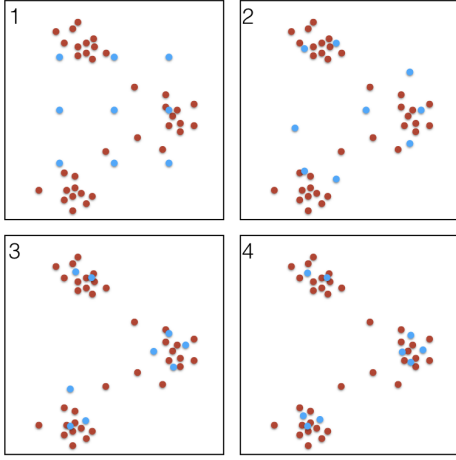


FIG. 1. Schematic illustration of the clustering problem considered in this paper. Blue dots represent clusters and red dots are data points. The training process moves clusters to fit the data points. Note that there are fewer clusters than data points which is the essence of dimensionality reduction and is what permits the model to generalize the data.

ing feedforward neural networks, one of the most usable neural network model, was developed [? ?]. Recent results connected classical Bayesian approach to deep learning allowed to develop a new algorithm for Bayesian deep learning on quantum computers [?]. The quantum models for convolutional neural networks, which may be suitable for the problems of learning of quantum states were also proposed [? ?]. Quantum classification, tested on the MNIST dataset, via Slow Feature Analysis based on the use of Quantum Frobenius Distance was proposed [?]. Sublinear quantum algorithms for training linear and kernel-based classifiers were designed [?]. Furthermore, continuous-variable quantum neural networks [?], quantum autoencoders [?], quantum approaches for reinforcement learning [? ? ?] and some other [? ? ? ? ? ? ? ?] were developed. Currently, it is believed that implementation of quantum neural networks can be the main test bed to achieve practical quantum supremacy on NISQ devices.

In this paper, we realize a hybrid quantum assisted SOFM (QASOFM) and apply it to the data clustering problem in an unsupervised manner on a toy example of clustering paper abstracts. We implement the quantum assisted SOFM in such a way that it becomes possible to use the Hamming distance as a distance metric for the training and to reduce the number of distance calculations in the number of clusters. We propose an optimized circuit for realizing the Hamming distance on a quantum machine. We prepare a toy example of data for clustering paper abstracts and give a proof-of-concept realization of the quantum assisted SOFM on the IBM Q experience quantum computer [?] and compare it with the classical case.

II. THE QUANTUM ASSISTED SELF-ORGANIZING FEATURE MAP (SOFM)

SOFM is one of the best-known unsupervised learning methods that is widely used in various areas of modern science. It was first proposed by Kohonen as a self-organizing unsupervised learning algorithm which produces feature maps similar to those occurring in the brain [?]. The SOFM algorithm operates with a set of input objects, each represented by a N -dimensional vector, as input and maps them onto nodes of a two-dimensional grid.

The input dimensions are associated with features and nodes in the grid, called cluster vectors, are assigned the N -dimensional vectors; the components of these vectors are usually called weights. Initially weight components are chosen randomly. We then can train our SOFM adjusting the components through the learning process which can be oversimplified into two basic procedures, selecting a winning cluster vector and updating its weights (Fig. ??). In more detail, they consist of four step process: 1) select an input vector randomly from the set of all input vectors; 2) find a cluster vector which is closest to the input vector; 3) adjust the weights of the winning node in such a way that it becomes even closer to the input vector; 4) repeat this process many iterations until it converges.

After the winning cluster vector is selected, the weights of the vector are adjusted according to

$$\vec{w}_{\text{new}} = \vec{w}_{\text{old}} + \alpha (\vec{x} - \vec{w}_{\text{old}}). \quad (1)$$

In simple words, this expression can be understood according to the following: if a component of the input vector is greater than the corresponding weight, increase the weight by a small amount; if the input component is smaller than the weight, decrease the weight by a small amount. The larger the difference between the input component and the weight component, the larger increment (decrement). Intuitively, this procedure can be geometrically interpreted as iteratively moving the cluster vectors in space one at a time in a way that ensures each move is following the current trends inferred from their distances to the input objects. A visualisation of this process is provided on figure (Fig. ??).

Usually the winning cluster vector is selected based on the Euclidean distance between an input vector and the cluster vectors. In our approach, we use the Hamming distance instead of the Euclidean distance to select the winning cluster vector. It allows us to use a simpler encoding of the classical information into the quantum state and use an effective procedure for the calculation of the Hamming distance on the quantum machine, such as to reduce the number of calculations in number of cluster vectors in comparison to the classical case.

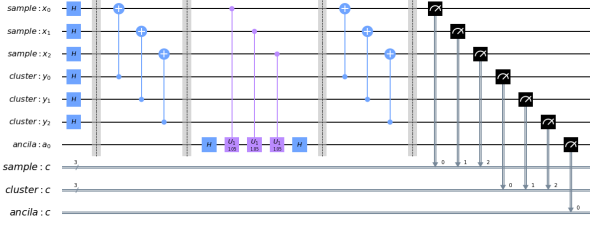


FIG. 2. A quantum circuit for the Hamming distance calculation algorithm between all samples and clusters at once. First, we apply Hadamard gates in order to get superposition of the sample and cluster registers. Second, we encoded information about pairwise different qubits in a quantum state of the sample register with applying the CNOT gates. Third, Hamming distance values are encoded in the amplitudes of superposition with the control phase rotation and Hadamard gates. Finally, a quantum state of the sample register returned to the initial basis for information retrieval.

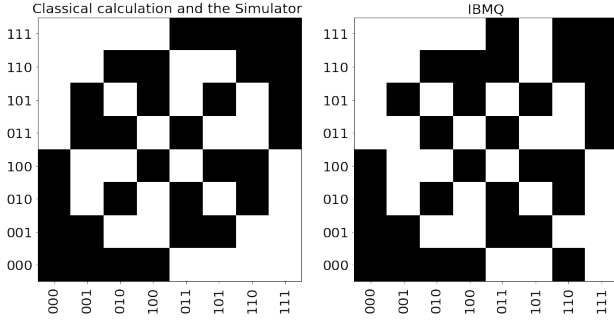


FIG. 3. The Hamming distance matrix between two data sets of binary vectors. Values of distance that less than median distance value was marked black. The classical simulation of the quantum circuit shows perfect agreement to theoretical calculations and presented on the left figure. Result obtained on IBMQ "ibmq_16_melbourne" backend is shown on the right figure. We can see good agreement between the distance matrices calculated classically and on the IBM Q Experience.

A. Optimized quantum scheme for Hamming distance calculation

We now introduce an optimized algorithm for calculating the matrix of Hamming distances [?] between a sample vector and all cluster vectors at once. This allows for a simple encoding of the classical information into a quantum register.

The overall procedure involves two registers of n qubits each, denoted $|X\rangle$ and $|Y\rangle$, along with a single auxiliary qubit $|a\rangle$. During entire processing, the $|Y\rangle$ is used to store the cluster states, then at the beginning and at the end of the processing the $|X\rangle$ register stores the input vectors, while during the processing it stores the differences between input vectors and cluster states.

Now for this demonstration let us assume we have k input vectors and l cluster states. The i th input vector and j th cluster vector are respectively denoted as $|x_i\rangle$

($|y_j\rangle$).

The registers $|X\rangle$ and $|Y\rangle$ will properly store input vectors and cluster vectors if initialized as follows.

$$|X\rangle = \frac{1}{\sqrt{k}} \sum_{i=1}^k |x_i\rangle, \quad (2)$$

$$|Y\rangle = \frac{1}{\sqrt{l}} \sum_{j=1}^l |y_j\rangle, \quad (3)$$

Both registers along with the auxiliary qubit compose into initial state of the entire circuit in a form denoted below.

$$|\psi_0\rangle = |X\rangle |Y\rangle |a\rangle \quad (4)$$

where $|a\rangle$ is an auxiliary qubit in the state $|0\rangle$ initially.

Given this initial state we may begin the processing of the problem. We start by applying a CNOT gate between $|X\rangle$ and $|Y\rangle$

$$|\psi_1\rangle = \text{CNOT}(Y, X) |\psi_0\rangle = \frac{1}{\sqrt{kl}} \sum_{i,j} |d_{ij}^1, \dots, d_{ij}^n\rangle |y_j^1, \dots, y_j^n\rangle |0\rangle \quad (5)$$

where $d_{ij}^\alpha = \text{CNOT}(y_i^\alpha, x_j^\alpha)$, $\alpha = 1 \dots n$ and i, j are the qubit indexes in the registers.

In the result, at this stage of the computation the $|X\rangle$ no longer stores the input vectors, instead it stores the information about pairwise different qubits between the input vector $\{X\}$ and cluster vector $\{Y\}$.

The next stage, for each pair $\{X\}$ and $\{Y\}$, an accumulated information of all the differences is projected into the amplitude of the superposed state. this is achieved by applying the Hadamard gate on auxiliary qubit, then control phase gate on $|Xa\rangle$,

$$R(\phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{-i\phi} \end{pmatrix}, \quad \phi = \frac{\pi}{n} \quad (6)$$

and again Hadamard gate.

After the first Hadamard on the ancillary qubit we have the following state

$$|\psi_2\rangle = H_a |\psi_1\rangle = \sum_{i,j} |d_{ij}^1, \dots, d_{ij}^n\rangle |y_j^1, \dots, y_j^n\rangle \frac{|0\rangle + |1\rangle}{\sqrt{2}} \quad (7)$$

Then we apply the control phase gate

$$\begin{aligned}
|\psi_3\rangle &= R_{(X,a)}\left(\frac{\pi}{n}\right)|\psi_2\rangle \\
&= \frac{1}{\sqrt{2}} \sum_{i,j}^k |d_{ij}^1, \dots, d_{ij}^n\rangle |y_j^1, \dots, y_j^n\rangle |0\rangle \\
&+ \frac{1}{\sqrt{2}} \sum_{i,j}^k \exp\left(\frac{-i\pi}{n} \sum_l^n d_{ij}^l\right) |d_{ij}^1, \dots, d_{ij}^n\rangle |y_j^1, \dots, y_j^n\rangle |1\rangle
\end{aligned} \tag{8}$$

Applying another Hadamard on ancillary qubit and doing necessary calculations the state of quantum register reads

$$\begin{aligned}
|\psi_4\rangle &= \sum_{i,j}^k e^{\left(\frac{-i\pi}{2n} \sum_l^n d_{ij}^l\right)} \\
&\times \left\{ \cos\left(\frac{\pi}{2n} \sum_l^n d_{ij}^l\right) |d_{ij}^1, \dots, d_{ij}^n\rangle |y_j^1, \dots, y_j^n\rangle |0\rangle \right. \\
&+ \left. i \sin\left(\frac{\pi}{2n} \sum_l^n d_{ij}^l\right) |d_{ij}^1, \dots, d_{ij}^n\rangle |y_j^1, \dots, y_j^n\rangle |1\rangle \right\} \tag{9}
\end{aligned}$$

This completed the stage of projecting the differences between pairs of $\{X\}$ and $\{Y\}$ onto the amplitude of the auxiliary qubit. The process is done in the x -basis, which explains the surrounding Hadamard gates. There are two possible measurement outcomes of the auxiliary qubit, each pair of $\{X\}$ and $\{Y\}$ forms a subspace of the Hilbert space, the controlled phase gate ensures to change amplitudes of those outcomes within this subspace depending on how different the spin configurations between $\{X\}$ and $\{Y\}$ are.

At this stage, the information of differences between pairs of $\{X\}$ and $\{Y\}$ is no longer relevant, thus we return to our initial basis for register $|X\rangle$ by applying the pairwise CNOT gates:

$$\begin{aligned}
|\psi_f\rangle &= \text{CNOT}(Y, X) |\psi_4\rangle \\
&= \sum_{i,j}^k e^{\left(\frac{-i\pi}{2n} \sum_l^n d_{ij}^l\right)} \left\{ \cos\left(\frac{\pi}{2n} \sum_l^n d_{ij}^l\right) |X_i\rangle |Y_j\rangle |0\rangle \right. \\
&\quad \left. + i \sin\left(\frac{\pi}{2n} \sum_l^n d_{ij}^l\right) |X_i\rangle |Y_j\rangle |1\rangle \right\} \tag{10}
\end{aligned}$$

This made $\{X\}$ store input vectors again, just like at the initial stage, however it preserved the amplitudes of the auxiliary qubit which are proportional to how different each pairs of $\{X\}$ and $\{Y\}$ are.

In the result we have the Hamming distances encoded into the amplitudes of the final state. After getting statistics of the measurement outcomes we can produce the

distance matrix between two data sets of binary vectors (Fig. ??). In this case, a biggest amplitude in measurement result coincide to a smallest Hamming distance when the measurement result of the auxiliary qubit is 0 and we have the inverse relationship when measurement result on the auxiliary qubit is 1.

Measuring the Hamming distance of a particular pair of input vector $|X_i\rangle$ and cluster vector $|Y_j\rangle$ consists of extracting the relevant amplitude from the subspace that those states form, this can be done using the following projector operator.

$$\Pi_{i,j} = |X_i\rangle\langle X_i| \otimes |X_i\rangle\langle Y_j| \otimes I \tag{11}$$

Using the above projection operator, the subspace of the Hilbert space formed by particular pair of input vector and cluster vector can be traced out as follows.

$$\rho_{i,j} = \text{Tr}_{1,\dots,2n}(\Pi_{i,j} |\psi_f\rangle\langle\psi_f| \Pi_{i,j}) \tag{12}$$

Now the following two relevant amplitudes for the measurement results can be extracted.

$$a_0(x_i, y_j) = \frac{\langle 0 | \rho_{i,j} | 0 \rangle}{\text{tr}(\rho_{i,j})}, \tag{13}$$

$$a_1(x_i, y_j) = \frac{\langle 1 | \rho_{i,j} | 1 \rangle}{\text{tr}(\rho_{i,j})}, \tag{14}$$

In order to reduce noise we average measurement results over different states of the auxiliary qubit, thus the measured Hamming distance between the input vector $|X_i\rangle$ and cluster vector $|Y_j\rangle$ takes the following form.

$$d_{i,j}^H \propto 1 - \frac{1}{2}(a_0(x_i, y_j) + (1 - a_1(x_i, y_j))) \tag{15}$$

The Hamming distance measured in this way is contained $0 \leq d_{i,j}^H \leq 1$, where intuitively, 0 would indicate x_i are y_j identical and 1 would mean they are completely opposite in terms of their pairwise binary coordinates.

In comparison with the circuits for Hamming distance calculations proposed previously [?], our method allows one to reduce number of gates in the circuit (at least in $2n$ one-qubit NOT gates, depends on actual realization control phase gate in a quantum register on the hardware level), reconstruct all distance matrix at once and can be implemented with high fidelity on current quantum devices.

III. DEMONSTRATION OF QASOFM ON THE IBM Q EXPERIENCE

We now show results for a proof-of-concept demonstration of the algorithm introduced in the previous section,

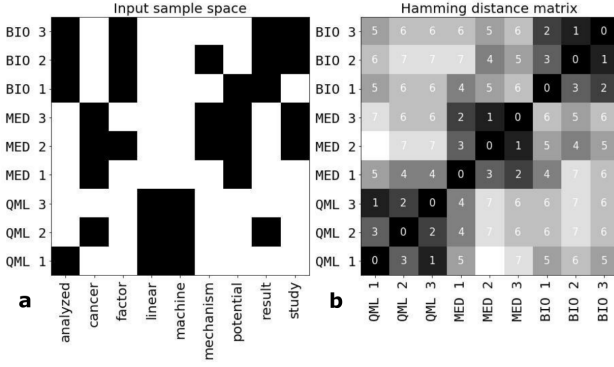


FIG. 4. (a) Representation of the data set of abstracts with the bag-of-words model is shown. Each abstract is represented by a binary vector with 9 elements. The samples are sorted into groups (QML, MED, BIO), 3 papers for each tag. (b) The Hamming distance between each vectorized abstract is shown with a number in the matrix. The sets of abstracts are well separated.

on a 16-qubit quantum computer provided by IBM Q Experience. We implement unsupervised data clustering for three sets of paper abstracts from three different fields: quantum physics [? ? ?], medicine [? ? ?] and biology [? ? ?]. Each set consists of three random papers that focus on one of following topics: “Quantum Machine Learning” (QML), “Cancer” (MED) and “Gene Expression” (BIO). Abstracts were vectorized by the bag-of-words model in order to choose most defining words in each data set (see Fig. ??). This model represents text as a multiset “bag” of its words taking into account only multiplicity of words. Preparing the bag-of-words we excluded the words that appear only in one abstract and more than in 4 abstracts and we also excluded the word “level” from consideration due to the frequent overlap between the clusters because it gives instabilities for both classical and quantum algorithms. We restricted our bag-of-word size to 9 of the most frequent words from the full “bags-of-word” due to limitations of the number of qubits. After vectorizing and pre-processing the data, the clusters are well separated with the Hamming distance. We observe that distances between the abstracts inside clusters are smaller than distances between the abstracts from different clusters, showing a successful self-organization (Fig. ??).

Classical SOFM [?], as it has been explained earlier, can be oversimplified into two basic procedures: distance calculation between a one sample vector and all cluster vectors and shifting the closest cluster vector to the sample vector. In classical SOFM, the complexity of algorithm in the sense of number of distance calculations scales as $O(LMN)$, where N is number of samples, M is number of randomly sampled cluster vectors, and L is number of the shifts of cluster vectors. In the QASOFM (Sec. ??), distance calculations are realized on a quantum device (i.e. the IBM Q Experience) with the use of circuit presented in Fig. ??. This approach allows one to

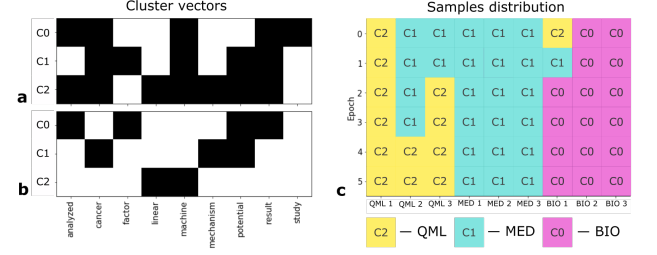


FIG. 5. (a) Initial random binary vectors of clusters are presented. We randomly sampled 3 binary vectors of size 9 as initial cluster vectors (labeled by 0,1,2) and it is shown how the words from the bag of words model presented in them. (b) The result of applying our QASOFM implemented on the IBMQ “ibmq_16_melbourne” backend is shown on the bottom left figure. Vectors mean cluster elements for BIO, MED, QML groups, respectively (up to down). (c) The evolution of label distribution on each learning epoch is presented on the right figure. Good convergence already is observed on the fourth epoch.

reduce the number of operations in a number of cluster states with an optimized number of gates that are possible to realize on currently available quantum computing devices with a limited number of qubits. The circuit realized in such a manner that calculation of Hamming distance between the sample vector and all cluster vectors is realized in one operation. The complexity of the quantum assisted SOFM then scales as $O(LN)$.

In order to check that our algorithm gives the correct results we compare it to classical calculations of the distance matrix on two data sets of binary vectors, as shown in Fig. ??. We see good agreement between the distance matrices calculated classically and on the IBM Q Experience. The theoretical calculations and classical simulations show perfect agreement with each other. An example of the QASOFM learning process is given on Fig. ??. Initially, the cluster vectors were randomly chosen (see Fig. ??) and label of sample distribution is shown for the zeroth epoch in Fig. ??(a). In order to prepare a superposition of cluster vectors needed for the calculation of the distance matrix at once we use the standard initialization of QISKIT library. Each epoch of the algorithm requires 9 distance calculations in the quantum implementation (number of samples in general case) or 27 distance calculations for classical realization (product of number of samples and number of cluster vectors in general case). In order to realize the circuit for the distance calculation for the quantum case we used the IBM Q machine. After the distance calculation from each sample to all cluster vectors at each epoch we label each sample with the index of the closest cluster vector and shift the closest cluster vectors to the sample one. The shift is made by the change of the first binary element in the cluster vectors different from the sample one. The evolution of the labels presented on Fig. ??(c). Good convergence is already observed in the fourth epoch.

IV. DISCUSSIONS

We have developed a quantum assisted SOFM and showed a proof-of-concept demonstration that it can be used to solve clustering problems in an unsupervised manner with potential to outperform classical counterpart in real life tasks. The procedure of solving such clustering problem requires calculating the distance many times in iterative way. We introduced an optimized circuit for Hamming distance calculations that can be implemented on currently available quantum computing devices with high fidelity. Our circuit could act as a distance-computing component of a classical SOFMs algorithm and in this way improve its performance. Due to wide use of classical SOFMs in different areas of modern research this can give opportunities for the use of QASOFM in practical applications in near term.

ACKNOWLEDGMENTS

The work is supported by the RFBR-NSFC collaborative program (Grant No. 18-57-53007) and the State assignment (N. 0089-2019-0002). I.D.L. acknowledges support from the Russian Foundation for Basic Research,

grant No.19-32-80004 and support from the Foundation for the Advancement of Theoretical Physics and Mathematics BASIS No.19-1-5-130-1. T.B. is supported by the Shanghai Research Challenge Fund; New York University Global Seed Grants for Collaborative Research; National Natural Science Foundation of China (61571301, D1210036A); the NSFC Research Fund for International Young Scientists (11650110425, 11850410426); NYU-ECNU Institute of Physics at NYU Shanghai; the Science and Technology Commission of Shanghai Municipality (17ZR1443600); the China Science and Technology Exchange Center (NGA-16-001); and the NSFC-RFBR Collaborative grant (81811530112).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

KEYWORDS

quantum self-organizing feature map; quantum machine learning; neural network; quantum unsupervised data clustering; Hamming distance.