A 64 bit quantum dragon data-set for machine learning

M.A. Novotny*,†, Yaroslav Koshka*,‡, G. Inkoom†, Vivek Dixit*

* Center for Computational Sciences, High Performance Computing Collaboratory, Mississippi State University, Mississippi State, Mississippi 39762-9627, USA
† Department of Physics and Astronomy, Mississippi State University, Mississippi State, Mississippi 39762-5167, USA
‡ Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, Mississippi 39762, USA

E-mail: man40@msstate.edu

Abstract. Design and examples of a sixty-four bit quantum dragon data-set are presented. A quantum dragon is a tight-binding model for a strongly disordered nanodevice, but when connected to appropriate semi-infinite leads has complete electron transmission for a finite interval of energies. The labeled data-set contains records which are quantum dragons, which are not quantum dragons, and which are indeterminate. The quantum dragon data-set is designed to be difficult for trained humans and machines to label a nanodevice with regard to its quantum dragon property. The 64 bit record length allows the data-set to be utilized in restricted Boltzmann machines which fit well onto the D-Wave 2000Q quantum annealer architecture.

1. Introduction

Machine learning is making inroads into many aspects of modern research and modern society [1], from social media algorithms to autonomous vehicle applications. Machine learning is also making inroads as an application used in computational [2, 3] and experimental science [4]. One type of machine learning methodology is restricted Boltzmann machines (RBM) [5]. The possibility of using quantum computers in machine learning has a recent review [6]. The application of quantum annealing machines to train and utilize RBM methods [7, 8] is well established. Here we note an RBM can be embedded onto the $K_{4,4}$ Chimera architecture of a D-Wave machine [9, 10, 11], and the current D-Wave 2000Q can embed an RBM with 64 hidden and 64 visible units (assuming no missing qubits). In order to test classical and quantum machine learning methodologies, a data set based on quantum, rather than classical, principles may be important.

Thus data-sets with 64 bit records which are based on quantum principles are useful to test current D-Wave architectures as applied to machine learning. The goal of this paper is to introduce the design of a labeled 64 bit data-set which is difficult for humans and RBMs embeddable on a D-Wave 2000Q to generalize from a training set to distinguish underlying patterns which are not in the training set. Because of the small record size, construction of such a data-set is not an easy task.
2. Quantum Dragon Nanodevices

The quantum dragon data-set (QDDS) we introduce is based on electron transport through nanodevices which have strong disorder, but when connected to appropriate semi-infinite leads have complete electron transmission, \( T(E) = 1 \) over a finite interval of energies. Quantum dragon devices with cylindrical symmetry were introduced in 2014 [12], including Bethe lattices and single-walled carbon nanotubes. Here we construct our QDDS based on a nanodevice with an underlying square lattice [13]. We use both open boundary conditions of two types. Free boundary conditions are denoted BC00 [14], and boundary conditions modified at the boundary, denoted BC-- [13]. Quantum dragon devices are found by performing an exact mapping for the transmission onto a one dimensional device with the same \( T(E) \). This can be accomplished when every slice \( j \) of the device is represented by an \( m \times m \) intra-slice matrix \( A_j \), and all slices have a common eigenvector \( \vec{v}_{\text{dragon}} \) with associated eigenvalue \( \lambda_j \). Table 1 shows the form for \( A_j \) and the common eigenvector \( \vec{v}_{\text{dragon}} \) for both boundary conditions in our QDDS.

Table 1. The two boundary conditions utilized for 64 bit quantum dragon data-sets. Within the nanodevice the on site energies are \( \epsilon_{i,j} \) and the intra-slice hopping strengths are \( t_{i,j;i+1,j} \).

| BC  | \( \vec{v}_{\text{dragon}} \) | \( A_j \) |
|-----|-------------------|------------------|
| BC00 | \( \frac{1}{\sqrt{2}} \begin{pmatrix} \sin \left( \frac{\pi}{4} \right) \\ \sin \left( \frac{\pi}{4} \right) \\ \sin \left( \frac{3\pi}{4} \right) \end{pmatrix} \) | \( \begin{pmatrix} \epsilon_{1,j} & -t_{1,j;2,j} & 0 \\ -t_{1,j;2,j} & \epsilon_{2,j} & -t_{2,j;3,j} \\ 0 & -t_{2,j;3,j} & \epsilon_{3,j} \end{pmatrix} \) |
| BC-- | \( \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \) | \( \begin{pmatrix} \epsilon_{1,j} & -t_{1,j;2,j} & 0 \\ -t_{1,j;2,j} & \epsilon_{2,j} & -t_{2,j;3,j} \\ 0 & -t_{2,j;3,j} & \epsilon_{3,j} - t_{2,j;3,j} \end{pmatrix} \) |

For our 64 bit QDDS we utilize a device with \( \ell = 4 \) slices, each with \( m = 3 \) atoms in the slice, as shown in Figs. 1 and 2. The hopping strength of the semi-infinite lead wires is chosen as the unit of energy. (Only two atoms in each semi-infinite lead wire are shown, as green spheres.) All inter-slice hopping strengths are also set to unity. The on site energy of the atoms in the lead wires is either chosen as the zero (setting the zero for energy), or if all slices have \( \vec{v}_{\text{dragon}} \) as a common eigenvector and all the associated eigenvalues \( \lambda_j = \lambda_{\text{dragon}} \) are the same, the on site lead energies are all set to \( \lambda_{\text{dragon}} \). For zero lead on site energy, one has propagation of electrons for energies in the interval \(-2 \leq E \leq 2\). In the figures, the radii of the red spheres are proportional to the on site energies \( \epsilon_{i,j} \), and the radii of the black (vertical) cylinders are proportional to the intra-slice hopping strengths \( t_{i,j;i+1,j} \). The green spheres are the lead atoms, and the cyan horizontal bonds all have strength unity.

We choose \( n_\# = 3 \) bits to represent both the intra-slice hopping strengths and the on site energies of the device. We take both the on site energies \( \epsilon_{i,j} \) and intra-slice hopping strengths \( t_{i,j;i',j} \) to be non-negative, and have the allowed values \( \{0,1,2,3,4,5,6,7\} \) for \( n_\# = 3 \). Thus in each of the \( \ell = 4 \) slices there are \( 2m - 1 = 5 \) parameters, represented by \( (2m - 1)n_\# = 15 \) bits. Our \( \ell = 4 \) devices thus are represented in 60 bits.

We use two additional bits to label the type of boundary condition, BC00 or BC--, leaving an additional bit to label other boundary conditions [13]. Two bits label the record. If the record is known to represent a quantum dragon, the nanodevice has \( T(E) = 1 \) for \(-2 \leq E \leq 2\), and is coded as 01. If the record representing the device is mappable (i.e. all \( A_j \) share a common eigenvector), but the eigenvectors have different eigenvalues for different \( j \), extending particular values to infinite \( \ell \) would lead to transmission \( T(E) \ll 1 \) due to Anderson localization, and hence the device is not a quantum dragon, and is coded as 10. If the mapping is not possible, the value for \( T(E) \) is such there may be some range of energies which have complete transmission.
Figure 1. Data set representatives for $\ell = 4$ and $m = 3$ lattice, attached to leads with boundary conditions BC00. The top two are quantum dragons, the middle two are mappable but not quantum dragons, and the last two are undecided.

In this case the device cannot be classified, and it is coded randomly as 00 or 11.

3. Quantum Dragon Data-Sets

We thus have 64 bit records in our QDDS, each representing a nanodevice together with its labels. For our 64 bit QDDS it is possible to calculate the number of possible records, the number of records which correspond to mappable devices, and the number of records corresponding to quantum dragons. These are shown in Table 2.

Table 2. The number of possible different nanodevices for quantum dragon data-sets for an $\ell=4$ and $m=3$ device with $n_\# = 3$.

| Type | label | # config | # mappable | # dragons |
|------|-------|----------|------------|-----------|
| BC00 | 00    | $2^{60} = 1.15 \times 10^{18}$ | $2^{24} = 1.68 \times 10^7$ | $2^2 = 64$ |
| BC-- | 01    | $(8 \times 36^2)^4 = 1.16 \times 10^{16}$ | $104^4 = 1.17 \times 10^8$ | 91208 |

For BC00 there are $2^{5\times4\times3} = 2^{60}$ different configurations, since there are $2m - 1 = 5$ tight binding parameters per slice, $\ell = 4$ slices, and $n_\# = 3$. Of these $2^{2\times4\times3} = 2^{24}$ are known to be mappable, since there are $\ell = 4$ slices and within each slice $j$ one needs $t_{1,j;2,j} = t_{2,j;3,j}$ for the hopping energy values and $\epsilon_{1,j} = \epsilon_{2,j} = \epsilon_{3,j}$ on site energy values. Of these mappable configurations, $2^{2\times3} = 64$ are quantum dragons since in all slices all $t_{i,j; i+1,j}$ values must be the same and all $\epsilon_{i,j}$ values must be the same. See Fig. 1 for representative configurations for BC00 boundary conditions. The BC00 dragon set is relatively easy for humans to determine whether or not the device is a quantum dragon, since there is a high degree of symmetry because every slice must be identical.

For BC-- one requires $t_{1,j;2,j} \leq \epsilon_{1,j}$ and $t_{m-1,j;m,j} \leq \epsilon_{m,j}$ so both the tight binding parameters and the elements of the $A_j$ matrices can be represented in $n_\#$ digits. For $n_\# = 3$ the calculation $\sum_{\epsilon=0}^{7} \sum_{t=\epsilon}^{7} 1 = 36$ gives the number of possible tight binding parameters for the boundary atoms and
their attached intra-slice hopping parameters, compared to the $2^{2n_\#} = 64$ representable values. Therefore since one requires for putting into bits $\epsilon_{1,j} \geq t_{1,j,2,j}$ and $\epsilon_{m,j} \geq t_{m-1,j,m,j}$ while all other $\epsilon_{i,j}$ can take $2^{n_\#}$ values, for $m = 3$ there are $(8 \times 36^2)^4$ possible records. A Mathematica program was written to test each of the $8 \times 36^2 = 10,368$ allowed values for $A_j$, and obtained 104 which had the appropriate vector $\vec{v}_{\text{dragon}}$ as an eigenvector with eigenvalue $\lambda_j$. Therefore, there are $104^4$ mappable configurations. Of these mappable configurations, the program gives for BC-- and $\ell = 4$, after taking into account degeneracies, a total of 91208 which are quantum dragons. See Fig. 2 for representative configurations for BC-- boundary conditions. As seen in Fig. 2, the BC-- QDDS is difficult for humans to utilize to identify quantum dragons, since the quantum dragon devices have no easily recognizable symmetry.

4. Conclusion and Discussion
A data-set with 64 bits was designed, because such a data-set can test a restricted Boltzmann machine which can be fit easily onto a D-Wave 2000Q quantum annealer [9, 10, 11]. The data set is built on a $m \times \ell$ square lattice nanodevice in the tight binding representation, with boundary conditions BC00 or BC--, and hence has $\ell(2m - 1)$ tight binding parameters. Each tight binding parameter is taken to be non-negative, and embedded in $n_\#$ bits. For our data-sets we choose $\ell = 4$, $m = 3$, and $n_\# = 3$, giving 60 bits to provide a record for a particular nanodevice. We use two additional bits to label whether or not the device is a quantum dragon or if the answer is unknown, as well as two bits to label the boundary condition used in the nanodevice.

Future data-sets can use different values for $\ell$, $m$, and $n_\#$ if larger restricted Boltzmann machines are embedable on a future quantum annealers. For the BC00 the number of total possible configurations are $\left(2^{(2m-1)n_\#}\right)^\ell$, the number of mappable configurations is $2^{2\ell n_\#}$, and the number of quantum dragon configurations is $2^{2n_\#}$. For BC-- the number of possible configurations is $\left[2^{(2m-5)n_\#} \times \left(\frac{2^{2n_\#} + 2^{n_\#}}{2}\right)^2\right]^\ell$, while the number of mappable and quantum dragon configurations can be obtained by writing a program to count the number, including taking into account symmetries.
The quantum dragon data-sets are very quantifiable. It is possible to calculate the total number of records, the number of mappable devices, and the number of quantum dragons. Thus a data set is comprised by choosing at random the three types of devices, each with a prescribed probability. Furthermore, some records can be left out of a training data-set, and left for a test data-set. The quantifiable nature of the data-set enables a measure of how well either a human or a machine learning paradigm can generalize from the training data-set to records which were not used in the training.

Acknowledgments

This material is based on research sponsored by the Air Force Research Laboratory (AFRL) under agreement number FA8750-18-1-0096. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory (AFRL) or the U.S. Government.

References

[1] Mohri, A., Rostamizadeh, A., & Talwakar, A. 2018 Foundations of Machine Learning, 2nd edition (Cambridge, MA, USA; MIT Press).
[2] Carleo, G. & Troyer, M. Science 355, 602-606 (2017).
[3] Tarelho, J.P.G., et al, Materials Today 21, 1019-1041 (2018).
[4] Wigley, P.B. et al. Scientific Reports 6, 25890 [6 pages] (2016).
[5] Aggarwal C.C. 2018 Neural Networks and Deep Learning: A Textbook (Cham, Switzerland; Springer International Publishing).
[6] Biamonte, J. et al. Nature 549, 195-202 (2017).
[7] Adachi, S.H. & Henderson, M.P. arXiv:1510.06356 (2015).
[8] Rasdi Rere, L.M. Fanany, M.I., & Arymurthy, A.M. Procedia Computer Science 72, 137-144 (2015).
[9] Koshka, Y., Perera, D., Hall, S., & Novotny, M.A. Proceedings of the International Joint Conference on Neural Networks, 1948-1954 (2016).
[10] Koshka, Y., Perera, D., Hall, S., & Novotny, M.A. Neural Computation 29, 1815-1837 (2017).
[11] Koshka, Y., Perera, D., Hall, S., & Novotny, M.A. Proceedings of the International Joint Conference on Neural Networks, 8489746 (2018).
[12] Novotny, M.A. Physical Review B 90, 165103 [14 pages] (2014).
[13] Inkoom, G. Ph.D. dissertation, Mississippi State University (2017).
[14] Inkoom, G. & Novotny, M.A. J. Physics Communications 2, 115109 [22 pages] (2018).