Training Restricted Boltzmann Machines Using a Quantum Annealer

Vaibhaw Kumar, Gideon Bass, and Joseph Dulny III
Booz Allen Hamilton
Washington DC, USA
kumar_vaibhaw@bah.com

***Abstract*—** Machine learning and the optimization involved therein is of critical importance for commercial and military applications. Due to the extremely complex nature of many-variable optimization, the conventional approach is to employ a meta-heuristic technique to find suboptimal solutions. Quantum Annealing (QA) hardware offers a completely novel approach for obtaining significantly better or optimal solutions with considerably large speed-ups when compared to traditional computing hardware. In this presentation, we describe our development of new machine learning algorithms tailored for QA hardware. We train a Restricted Boltzmann Machine (RBM) using QA hardware as a sampler. We present our initial results obtained by training RBMs on an image data set. We also discuss strategies for scaling up, including enhanced embedding and partitioned RBMs, to overcome the limitation imposed by current QA hardware.

Keywords—machine learning, deep learning, quantum computing, Boltzmann distribution, sampling methods, quantum annealing

I. INTRODUCTION

Deep learning techniques are relatively new in the field of machine learning [1]. These techniques are based on the premise that the input data in tasks such as image and speech recognition contain multiple layers of abstraction. For example, in an image of a player playing soccer, individual pixelated objects such as ball and player each form a separate layer of abstraction. Several such layers of abstraction give rise to a complete image. In order to model this data in any meaningful manner, one requires a good understanding of the abstraction present in the data. Traditional rule-based machine learning algorithms require extremely tedious and nontrivial manual feature engineering to perform well. Deep learning algorithms automatically perform such feature engineering for the user and automatically learn several levels of abstraction present in the input data.

The restricted Boltzmann machine (RBM) is a popular technique which learns the underlying population distribution of the input data using a simplified neural network. Stacked RBMs (deep belief networks) give rise to a deep learning architecture which can capture multilayer abstraction present in the data in a highly robust manner. They can be used both as a powerful autoenconder and as a classifier for machine learning tasks. They find applications in image and text recognition, anomaly detection, collaborative filtering and dimensionality reduction [2], [3].

II. RESTRICTED BOLTZMANN MACHINES

An RBM consists of a two-layered network where the stochastic binary input data is connected to stochastic feature detectors via symmetrically connected weights. The RBM is an energy-based model with a model Hamilton that bears a close resemblance to quadratic unconstrained binary optimization (QUBO). In order for the model to learn the data distribution, the log-likelihood of the data is minimized [2]. The "model average" term in the minimization is computationally difficult to compute and current techniques such as k-step contrastive divergence (CD-k) [2] and persistent contrastive divergence (PCD) [4] can lead to poor learning results. Additionally, the representations created by RBMs trained with these techniques can be extremely inaccurate [3]. Thus, techniques for improved RBM learning are needed.

III. QUANTUM ANNEALED SAMPLING

Recent studies [6] suggest that QA hardware can be used as a sampler to generate Boltzmann and Boltzmann-like distributed samples. The role of the QA hardware as a sampler is to provide an extremely efficient and accurate way to compute the "model averages" terms during the log-likelihood minimization of the RBM. Several studies have been conducted along this line of research: Dumoulin et al. [7] have extensively studied the influence of the inherent limitations of the QA hardware such as limited connectivity and parameter noise on the performance of an RBM using a quantum annealing simulator. Adachi and Henderson [8] have successfully implemented a hardware graph-restricted RBM on the D-Wave QA chip by representing each logical variable as a string of quantum bits (qubits) connected by ferromagnetic couplings. They use RBMs for pre-training a classifier for an image classification task. More recently, Benedetti et al. [6] have carried out a careful study on the use of QA hardware as a sampler for RBMs. This study used a synthetic "Bars and Stripes" dataset and described how to extract effective temperature from the sampled data and how to use temperature rescaling to improve the accuracy of RBMs.

IV. METHODS

A. *Techniques*

Most of the studies so far have worked with moderate-size datasets and were primarily focused on the implementation details of RBMs. Any work involving actual implementation on a relatively large commercial dataset is still missing. This is primarily due to several restrictions imposed by the QA hardware.

In this work, we develop strategies which allow us to circumvent the restrictions imposed by modern QA hardware. These strategies improve the accuracy as well as the capacity of RBMs to handle large, high-dimensional datasets. We accomplish this through two different research directions:

1) **Using generalized Boltzmann machines for higher connectivity**: We are developing a different embedding scheme in order to improve the connectivity between the network units of the RBMs. This will be implemented on QA hardware by creating ferromagnetic chains of qubits. While increasing connectivity, creating qubit chains can limit the number of logical qubits available on the chip. We plan to circumvent this issue by utilizing a partitioned RBM technique. We also plan to implement generalized network architecture such as semi-restricted Boltzmann machines and deep networks such as Deep Belief networks/Deep Boltzmann machines on QA hardware.

2) **Utilizing partitioned "atomic" RBMs for high-dimensional data**: We propose a partitioned RBM technique [9] which will allow us to efficiently work with high-dimensional large datasets by dividing the feature space into small subsets. Several small "atomic" RBMs will work on a given subset of the feature space. Due to the small size of the individual RBM, it can be completely embedded on the QA hardware and enable us to work with the limited number of qubits present on the hardware chip.

V. RESULTS

Taking our initial steps in this direction we have modelled two systems: 1) an Ising model containing six spins and 2) a 4x4 pixel "Bars and Stripes" image dataset. We show here results for an RBM trained on Ising model data. Ising model data set was artificially created using MCMC simulations. A QA simulator was used to train an RBM on the dataset. Fig.1 indicates a comparison between an RBM trained using QA simulator and CD-1 method. Our initial results indicate that the learning based on QA simulator is much better compared to heuristic method such as CD-k. Additional results using the techniques outlined above are currently under development.

VI. CONCLUSIONS AND FUTURE WORK

The techniques discussed in this work will allow us to successfully map real-life, large, and high-dimensional datasets onto the QA platform. A systematic study using the proposed methodology in already underway.

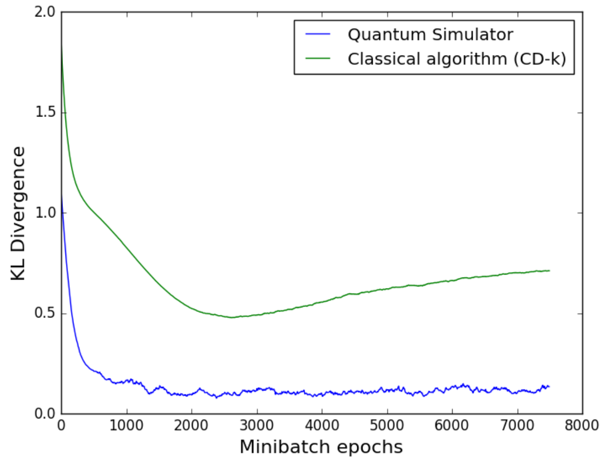

Fig. 1. The evolution of KL Divergence with the number of training epochs. Green and blue lines indicate CD-1 and a QA simulated RBMs, respectively.

REFERENCES

[1] Y. Bengio, *Learning Deep Architectures for AI*, vol. 2, no. 1. 2009.

[2] G. E. Hinton, "Reducing the Dimensionality of Data with Neural Networks," *Science (80-.).*, vol. 313, no. 5786, pp. 504–507, 2006.

[3] H. Larochelle and Y. Bengio, "Classification using discriminative restricted Boltzmann machines," *Icml*, pp. 536–543, 2008.

[4] T. Tieleman, "Training Restricted Boltzmann Machines using Approximations to the Likelihood Gradient," *Proc. 25th Int. Conf. Mach. Learn.*, vol. 307, p. 7, 2008.

[5] A. Fischer and C. Igel, "An Introduction to Restricted Boltzmann Machines," *Lect. Notes Comput. Sci. Prog. Pattern Recognition, Image Anal. Comput. Vision, Appl.*, vol. 7441, pp. 14–36, 2012.

[6] M. Benedetti, J. Realpe-Gómez, R. Biswas, and A. Perdomo-Ortiz, "Estimation of effective temperatures in a quantum annealer and its impact in sampling applications: A case study towards deep learning applications," *arXiv:1510.07611*, no. Mcmc, pp. 1–13, 2015.

[7] V. Dumoulin and I. Goodfellow, "On the Challenges of Physical Implementations of RBMs," *arXiv Prepr. arXiv …*, pp. 1199–1205, 2013.

[8] S. H. Adachi and M. P. Henderson, "Application of Quantum Annealing to Training of Deep Neural Networks," *arXiv:1510.00635*, no. Cd, pp. 1–18, 2015.

[9] H. Tosun, "Efficient Machine Learning Using Partitioned Restricted Boltzmann Machines," 2016.