RBM Image Generation Using the D-Wave 2000Q

Jennifer Sleeman, Milton Halem, John Dorband
University of Maryland, Baltimore County

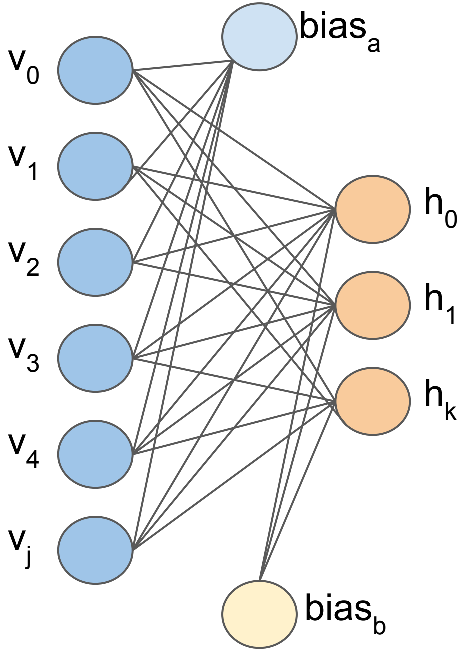
Presented at the D-Wave Qubits North America Conference
September 2019

Introduction

In this work, we describe a hybrid system that uses a deep convolutional neural network autoencoder and a quantum Restricted Boltzmann Machine (RBM) for image generation using the D-Wave 2000Q

Background - Restricted Boltzmann Machine

- A simple neural network used to find stochastic representations of the input
- Probabilistic, graphical model
- One visible layer $v_0 \ldots v_j$
- One hidden layer $h_0 \ldots h_k$
- Bias vectors a and b
- Restricted: No connections between nodes within a layer



Background - Restricted Boltzmann Machine

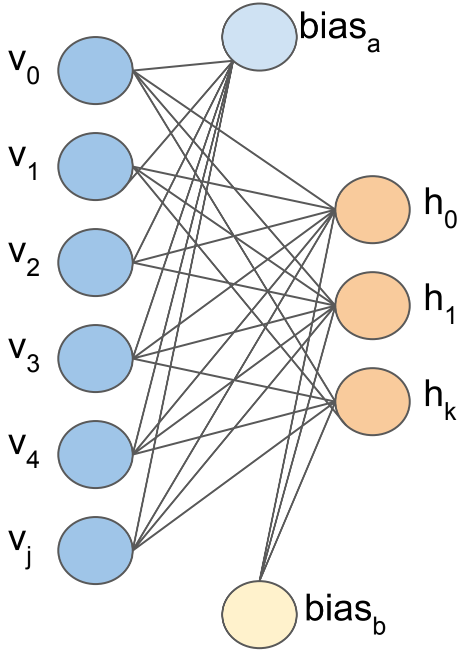
Represents a probability distribution over the visible (***v***) and hidden (***h***) units

Uses an energy-based function ***E*** for measuring quality of the model, minimizes:

$$E(\mathbf{v},\mathbf{h}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{i,j} v_i h_j w_{ij}$$

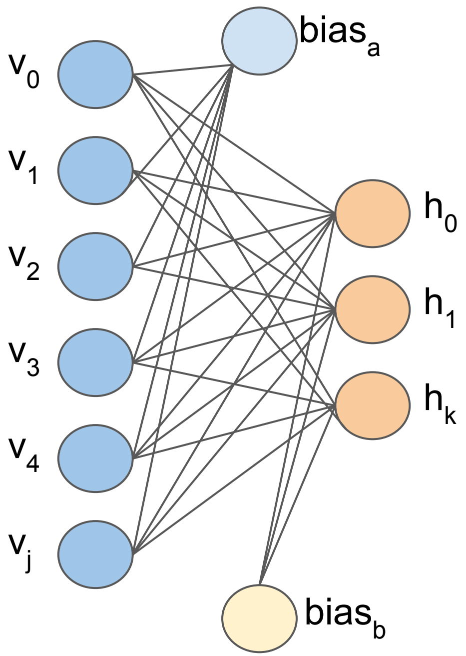
Joint probability mass function represented as:

$$p(\mathbf{v},\mathbf{h}) = \frac{1}{Z} e^{-E(\mathbf{v},\mathbf{h})} \qquad Z = \sum_{\mathbf{v},\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h})}$$



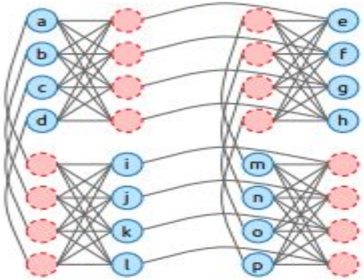
Background - Restricted Boltzmann Machine

- Gradient calculations of joint probability distribution are intractable
- Gibb's sampling is used to sample from the joint Boltzmann distribution using Markov Chain Monte Carlo (MCMC)



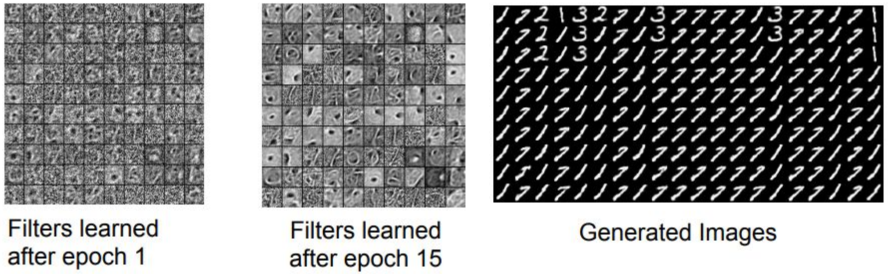
Background - Restricted Boltzmann Machine

- D-Wave connectivity 2-D grid, can be constrained to be chimera graph
- Chimera graph can be partitioned to form bipartite graph
- For image processing we formulate an embedding based on the number of pixels of the image (We use the MNIST dataset) using an RBM based on [15].



a	b	e	f
c	d	g	h
i	j	m	n
k	l	o	p

In [11] - Chimera-RBM embedding - mapping of pixels to visible units



Based on previous work [10] - we used D-Wave for sampling

What is a Generative Model?

Generative models model the joint probability distribution rather than a conditional probability distribution

For image generation, deep networks are used to learn a distribution that is similar to the true data distribution

The distribution of sampled output does not have a relationship with the distribution of samples from input variables

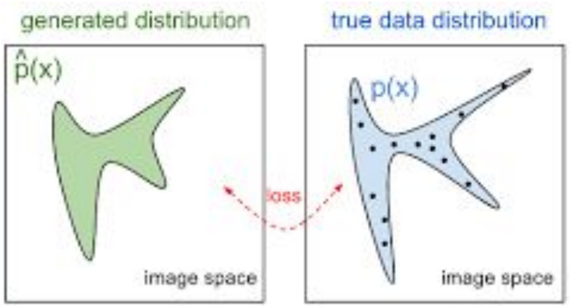


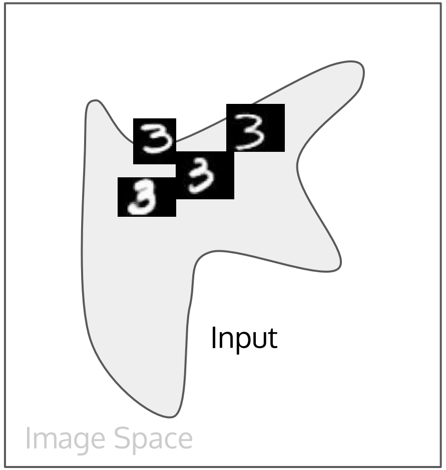
Image credit: OpenAI

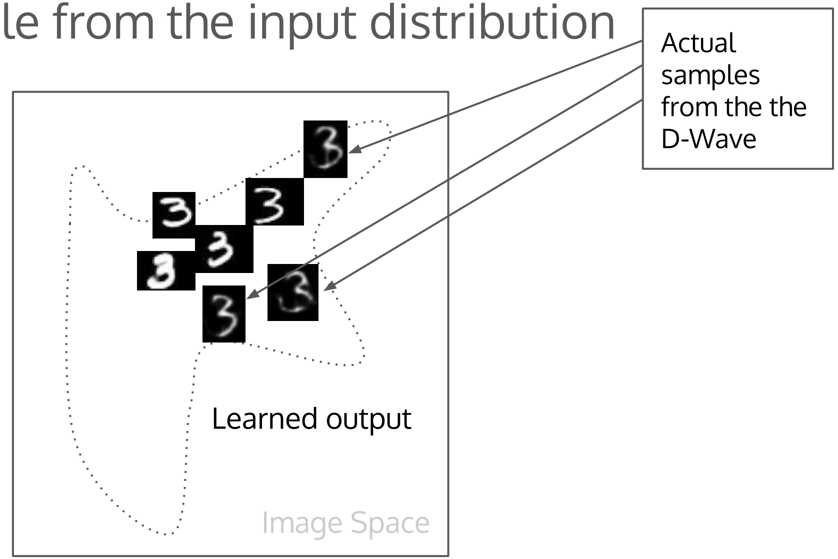
The Generative Properties of an RBM

RBMs learn a joint probability distribution $P(v,h)$

Where v represents the visible units and h represents the hidden units

Given $P(v,h)$, sampling from this distribution, *could* enable generated output that is not necessarily a replication of a sample from the input distribution





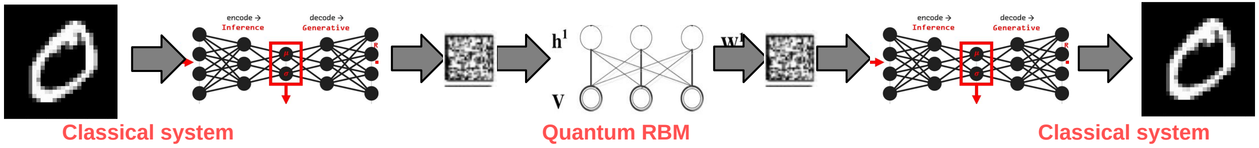
Challenges Using the D-Wave for Generative Sampling

Using the D-Wave for RBM image sampling is challenged by three main issues:

- Qubit size limitations makes it harder to embed large problems
- Samples from the D-Wave are binary and need to be mapped to their floating point values
- Generative sampling using a similar approach for embedding the RBM in the past has yielded poor results [10]

A Hybrid Approach

- To overcome these limitations,we mapped the original image space to a compressed image space.
- We compress MNIST digits from 28 x 28 to 7 x 7 and also 28 x 28 to 6 x 6
- We use the D-Wave API for working with the quantum annealer
- These experiments do not include the new D-Wave Hybrid API



Results

The original grayscale digits, even at a compression of 6 x 6 in size, is recoverable using our method when evaluating just the translation method in isolation

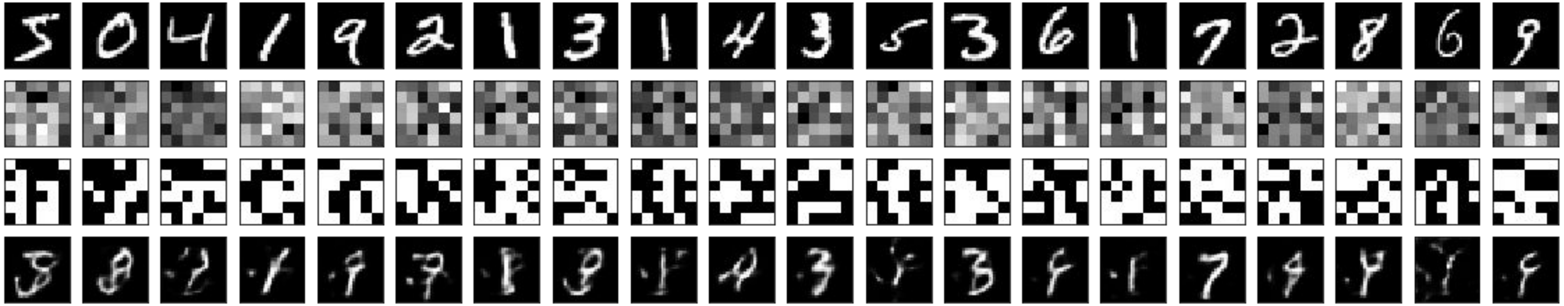
Shown are MNIST digits recovered after going through the translation process

28x28 Grayscale MNIST

36 bit Encoded Information

36 bit Binary Encoded Information

28 x 28 Recovered MNIST Digits



Translation from MNIST grayscale images, to encoded representation, to binary embedding, to grayscale

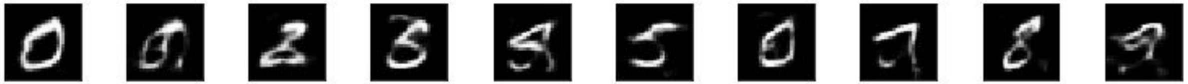
Results

MNIST digits recovered after training the RBM using the D-Wave and going through the translation process after sampling digits 0-9

Digits Represented	0	1	2	3	4	5	6	7	8	9

D-Wave 7 x 7 sampled 28 x 28 Recovered MNIST Digits

D-Wave 6 x6 sampled 28 x 28 Recovered MNIST Digits

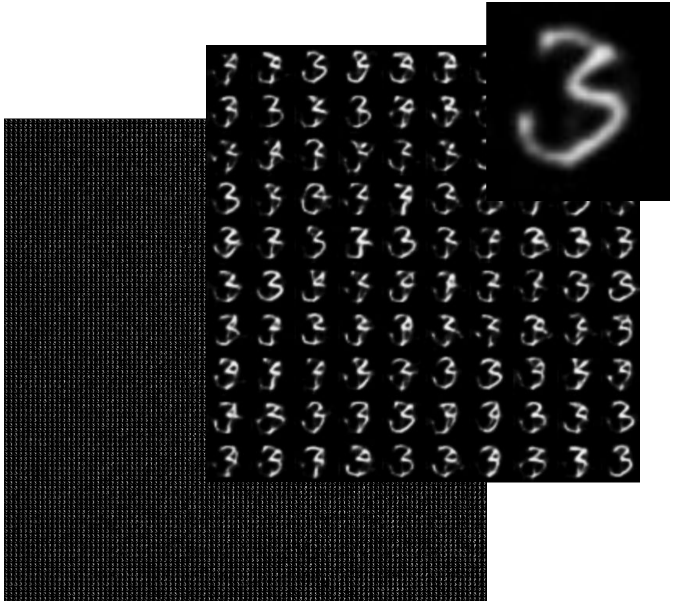


Results

Using the MNIST digit *3* - we show a large sampling, of this digit after training the RBM using the D-Wave

Using an RBM and the D-Wave as a generative sampler, we believe we can generate *NEW* variations of MNIST digits

We haven't proven this yet.

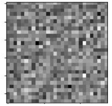


Early Results

Using a downstream process, we formulate a MNIST classification problem, whereby we use samples from our methods to train the MNIST classifier and compare the test results against an MNIST classifier with the original MNIST training data set. Training data consist of 60,000 samples, test set consist of 10,000 samples. Trained for 5 epoch only. Using a convolutional neural network.

Training Data Variants	***Accuracy On a Test set***	***Validation Loss***
Original MNIST 28 x 28 digits	0.9919	0.025
Sampled from Classical RBM MNIST 28 x 28 digits (no encodings)	0.9784	0.0912
Encoded/Decoded Binary Translated MNIST 28 x 28 digits (no RBM)	0.9717	0.232
Encoded/Decoded Binary MNIST 16 x 16 to translated to 28 x 28 digits (no RBM)	0.9534	0.195
Encoded/Decoded Binary MNIST 6 x 6 to translated to 28 x 28 digits (no RBM)	0.909	0.374
Encoded/Decoded Binary MNIST 6 x 6 to translated to 28 x 28 digits Sampled from the RBM D-Wave	0.6062	3.938

Note: Training with a classical RBM using both 28 x 28 and 6 x 6 binary encodings yielded results that could not be translated to a digit - ongoing.

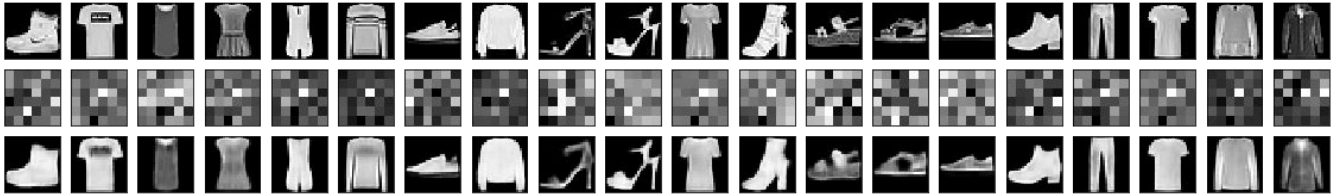


Classical RBM sample

Ongoing Experiments

We are currently running the following experiments:

1. Comparing performance and accuracy of classical and quantum RBMs
2. Evaluating the generative samples from the D-Wave against the original MNIST data and against the classical RBM samples
3. How do different hyperparameters affect/improve accuracy
4. Can we use other datasets?



Fashion-MNIST

Conclusions

This work puts forth a way to overcome qubit size limitations on the D-Wave by training on an embedded representation of images

By using a hybrid classical to quantum translation, we show promising results that indicate we may be able to use the D-Wave for image generation

Still a lot of work to be done to prove this is true

Thanks to:

References

[1] - Amin, Mohammad H., et al. "Quantum boltzmann machine." *Physical Review X* 8.2 (2018): 021050.

[2] - Kulchytskyy, Bohdan, et al. "Quantum Boltzmann Machine." *APS Meeting Abstracts*. 2016.

[3] - Adachi, Steven H., and Maxwell P. Henderson. "Application of quantum annealing to training of deep neural networks." *arXiv preprint arXiv:1510.06356* (2015).

[4] - Biamonte, Jacob, et al. "Quantum machine learning." *Nature*549.7671 (2017): 195.

[5] - Dorband, John E. "A boltzmann machine implementation for the d-wave." *Information Technology-New Generations (ITNG), 2015 12th International Conference on*. IEEE, 2015.

[6] - Khoshaman, Amir, et al. "Quantum Variational Autoencoder." *arXiv preprint arXiv:1802.05779* (2018).

[7] - Romero, Jonathan, Jonathan P. Olson, and Alan Aspuru-Guzik. "Quantum autoencoders for efficient compression of quantum data." *Quantum Science and Technology* 2.4 (2017): 045001.

[8] - Simonovsky, Martin, and Nikos Komodakis. "GraphVAE: Towards Generation of Small Graphs Using Variational Autoencoders." *arXiv preprint arXiv:1802.03480* (2018).

[9] - Sato, Issei, et al. "Quantum annealing for variational Bayes inference." *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. AUAI Press, 2009.
Hashizume, Yoichiro, et al. "Singular-value decomposition using quantum annealing." *Physical Review E* 92.2 (2015): 023302.

[10] - Thulasidasan, Sunil. *Generative Modeling for Machine Learning on the D-Wave*. LA-UR-16-28813, 1–23, 2016.

[11] - Benedetti, Marcello, et al. "Estimation of effective temperatures in quantum annealers for sampling applications: A case study with possible applications in deep learning." Physical Review A 94.2 (2016): 022308.

[12] - Dumoulin, Vincent, et al. "On the Challenges of Physical Implementations of RBMs." AAAI. Vol. 2014. 2014.

[13] - Rolfe, Jason Tyler. "Discrete variational autoencoders." arXiv preprint arXiv:1609.02200 (2016).

[14] - Hashizume, Yoichiro, et al. "Singular-value decomposition using quantum annealing." Physical Review E 92.2 (2015): 023302.

[15] - Ni and Nagayama, "Performance comparison on CFRBM between GPU and Quantum Annealing", Sept 2018

Questions?