A quantum-assisted algorithm for sampling applications in machine learning

Alejandro Perdomo-Ortiz
Research Scientist, Quantum Artificial Intelligence Laboratory
NASA Ames Research Center, Moffett Field, Calif., USA

Collaborators:
M. Benedetti, J. Realpe-Gomez, and R. Biswas.


Funding:
Unsupervised learning (generative models)

Example application: Image reconstruction

Reconstructed image

LEARNED MODEL

\( P(\text{Image}) \)

Damaged image
Unsupervised learning (generative models)

Learn the “best” model distribution that can generate the same kind of data.

Example application:
Image reconstruction

MODEL

\[ P(\text{Image}) \]

Learning algorithm

NO LABELS

DATASET

Example application:
Image reconstruction

LEARNED MODEL

\[ P(\text{Image}) \]

Reconstructed image

Damaged image

3
Supervised learning (discriminative models)

Learn the “best” model that can perform a specific task

MODEL
\[ P \left( \text{Label} \mid \text{Image} \right) \]

Learning algorithm

Example application: Image recognition

Predicted label

LEARNED MODEL
\[ P \left( \text{Label} \mid \text{Image} \right) \]

Image to be recognized

DATASET

26624 98 66 175

Labels
Outline

• Why is it hard and interesting to sample from a Boltzmann distribution? Why, in principle, is it possible to do classical Gibbs sampling with a quantum annealer?

• How to do it experimentally? Results on our quantum-assisted learning (QuALe) algorithm for sampling applications. Feasibility question.

• Overcoming the “curse of limited connectivity” in hardware. How to work with general probabilistic graphical models beyond RBM? How to cope with noisy devices and future directions.

Amin. PRA, 92, 052323 (2015)

Unsupervised learning relies on sampling

“Unsupervised learning [...] has] been overshadowed by the successes of purely supervised learning. [...] We] expect unsupervised learning to become far more important in the longer term. Human and animal learning is largely unsupervised: we discover the structure of the world by observing it, not by being told the name of every object.”


“In the context of the deep learning approach to undirected modeling, it is rare to use any approach other than Gibbs sampling. Improved sampling techniques are one possible research frontier.”


“Most of the previous work in generative models has focused on variants of Boltzmann Machines [...] While these models are very powerful, each iteration of training requires a computationally costly step of MCMC to approximate derivatives of an intractable partition function (normalization constant), making it difficult to scale them to large datasets.”

Mansimov, Parisotto, Ba, Salakhutdinov, under review for ICLR 2016
Restricted Boltzmann Machines and Beyond

**Deep Belief Networks**

**Training Method:** Stochastic gradient ascent

\[
\sum_{\mathbf{v} \in S} \frac{\partial \ln \mathcal{L}(\theta | \mathbf{v})}{\partial W_{ij}} \propto \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}}
\]

\[
= \langle v_i h_j \rangle_{p(h|v)q(v)} - \langle v_i h_j \rangle_{p(h,v)}
\]

**RBM’s:**

\[
E(\mathbf{v}, \mathbf{h}) = - \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} h_i v_j - \sum_{j=1}^{m} b_j v_j - \sum_{i=1}^{n} c_i h_i
\]

such that

\[
p(h|v) = \prod_{i=1}^{n} p(h_i|v) \quad \text{and} \quad p(v|h) = \prod_{j=1}^{m} p(v_j|h).
\]

**Model:**

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

Computationally bottleneck
**Foundational Theory of Quantum Annealing**

**Simulated Annealing**  
(Kirkpatrick et al., 1983)

- **Algorithm:** Start with a high temperature. Slowly reduce the intensity of these thermal fluctuations aiming for low cost configs.
- Transitions between states are over the barrier and due to thermal fluctuation

**Quantum Annealing**  
(Finnila et al., 1994, Kadowaki and Nishimori, 1998, Farhi et al., 2001)

- **Algorithm:** Start with large amplitude $A(\tau)$ responsible for quantum fluctuations. Then, slowly turn it off while turning on the cost function amplitude, $B(\tau)$.
- Transitions between states due to quantum fluctuations (tunneling)

$$H(\tau) = A(\tau) H_b + B(\tau) H_p$$

$$H_p = \sum_{1 \leq i \leq N} h_i \sigma_i^z + \sum_{1 \leq i < j \leq N} J_{ij} \sigma_i^z \sigma_j^z$$

$E(\{z\})$: Free energy Surface (cost funct.)

$\{z\}$=configurations in solutions space

- Initialize in an easy to prepare full quantum superposition
- Quantum states explored by quantum tunneling
- Final states: bit strings encoding the solution.
D-Wave System Capability

1) As a discrete optimization solver:

Given \{ h_j, J_{ij} \}, find \{ s_k = \pm 1 \} that minimizes the expression \( \xi(s_1, \ldots, s_N) = \sum_{j=1}^{N} h_j s_j + \sum_{i,j \in E} J_{ij} s_i s_j \)

Potential NASA applications: planning, scheduling, fault diagnosis, graph analysis, communication networks, etc.

Also, quantum ML work by Google/DW.

QUBO: Quadratic Unconstrained Binary Optimization (Ising model in physics jargon).

2) As a physical device to sample from Boltzmann distribution:

\[ P_{\text{Boltzmann}} \propto \exp \left[ -\frac{\xi(s_1, \ldots, s_N)}{T_{\text{eff}}} \right] \]

Potential NASA applications in machine learning (e.g., training of deep-learning networks)


Our recent work: Benedetti et al. PRA, 94, 022308 (2015)

- We provide a robust algorithm to estimate the effective temperature of problem instances in quantum annealers.
- Algorithm uses the same samples that will be used for the estimation of the gradient
Why sampling from classical Gibbs?

2) As a physical device to sample from Boltzmann distribution:

$$P_{\text{Boltzmann}} \propto \exp \left[ -\frac{\xi(s_1, \ldots, s_N)}{T_{\text{eff}}} \right]$$

Potential NASA applications in machine leaning (e.g., training of deep-learning networks)
Quantum-Assisted Learning Vs. Contrastive Divergence

Bars and Stripes dataset

Fisher and Igel. *Pattern Recognition, 47, 25 (2014)*

Embedding on the D-Wave 2X

Benedetti et al. PRA, 94, 022308
Non-trivial and correlated variations in the temperature

\[ T_{\text{DW2X}} = 0.033 \]

Benedetti et al. PRA, 94, 022308
Added features: Restart from CD-k

\[ L_{av} \]

-6
-7
-8
-9
-10
-11
-12

0 100 200 300 400 500

iteration

CD-1
QuALe with restart @ 250
QuALe @ \( T = 0.033 \) with restart @ 250

Benedetti et al. PRA, 94, 022308
Comparison with pseudo-likelihood

(a)

Comparison of QuALe with pseudo-likelihood and linear regression over iterations. The graph shows the average log-likelihood ($L_{av}$) over iterations for both methods. The legend indicates the different methods:

- QuALe w/ pseudo likelihood
- QuALe w/ linear regression

The graph is from Benedetti et al. PRA, 94, 022308.
Quantum-assisted unsupervised learning on digits

Overcoming the curse of limited connectivity

7 logical (visible) variables

18 physical qubits
Quantum-assisted unsupervised learning on digits

Overcoming the curse of limited connectivity in physical devices.

42 fully-connected logical (visible) variables

How do we train this 794 qubit problem? (How do we analyze the (Gibbs) samples from this physical model?)

Immediate solution: Keep an eye on a paper coming out with a new gray-model approach for training noisy QA.

Quantum-assisted unsupervised learning on digits

OptDigits Datasets

Dataset: Optical Recognition of Handwritten Digits (OptDigits)
Quantum-assisted unsupervised learning on digits

OptDigits Datasets

Dataset: Optical Recognition of Handwritten Digits (OptDigits)
Quantum-assisted unsupervised learning on digits

Dataset: Optical Recognition of Handwritten Digits (OptDigits)

Quantum-assisted unsupervised learning on digits

- Experimental realization of quantum-assisted learning algorithm on 794 qubits, for a 42 fully-connected model.
- Fully unsupervised learning and generative model on a digit.

Dataset: Optical Recognition of Handwritten Digits (OptDigits)

Quantum-assisted unsupervised learning on digits

- Experimental realization of quantum-assisted learning algorithm on 794 qubits, for a 42 fully-connected model.
- Fully unsupervised learning and generative model on a digit.

Dataset: Optical Recognition of Handwritten Digits (OptDigits)

Quantum-assisted unsupervised learning on digits

- Experimental realization of quantum-assisted learning algorithm on 794 qubits, for a 42 fully-connected model.
- Fully unsupervised learning and generative model on a digit.

Dataset: Optical Recognition of Handwritten Digits (OptDigits)

Quantum-assisted unsupervised learning on digits

- Experimental realization of quantum-assisted learning algorithm on 794 qubits, for a 42 fully-connected model.
- Fully unsupervised learning and generative model on a digit.

Dataset: Optical Recognition of Handwritten Digits (OptDigits)

Quantum-assisted unsupervised learning on digits

Overcoming the curse of limited connectivity in hardware.

46 fully-connected logical (visible) variables

42 for pixels + 4 to one-hot encode the class (only digits 1-4)

Are the results from this training on 917 qubit experiment meaningful? Is the model capable of generating digits, as expected?
Quantum-assisted unsupervised learning on digits

Human or (quantum) machine? (Turing test)

Dataset: Optical Recognition of Handwritten Digits (OptDigits)

Quantum-assisted unsupervised learning on digits

Human or (quantum) machine? (Turing test)

• Experimental realization of quantum-assisted learning algorithm on 917 qubits, for a 46 fully-connected model.

Dataset: Optical Recognition of Handwritten Digits (OptDigits)

Quantum-assisted unsupervised: artificial model

Reference model: Ising spin glass with 20 fully-connected spins (10 instances).

Ongoing research directions

Possible further boosting protocols by considering models to account explicitly for the noise in the quantum device.

Numerical simulations show that main limitation of current quantum annealers for Boltzmann machines applications is its sparse connectivity.

Extensions to deep learning architectures.

How “Boltzmannian” need the samples to be for QuALE to work.

Inference by using quantum distributions, such as those coming from future generation quantum computing technologies.

Is quantum tunneling, or any other quantum computational resource, relevant for machine learning/sampling applications? Can it be any faster than MCMC? Is it possible to achieve quantum supremacy in this domain?
Support slides