Experimenting with machine learning algorithms on quantum computers

Nagy Dániel[1]

[1]Institute for Physics, Eötvös Loránd University, H-1117, Pázmány Péter sétány 1/A. Budapest, Hungary

Created on December 1, 2019

Last update: March 6, 2021

Abstract

A modern számítástechnika jelentős eredményei közé tartozik a gépi tanulás és mesterséges intelligencia alapvető algoritmusainak kifejlesztése és ezek hasznosságának tesztelése különböző feladatokon. Ugyanakkor az elmúlt években a kvantumszámítás is jelentős fejlődésen ment keresztül, olyannyira, hogy 2019-ben a Google kísérleti csapatának sikerült demonstrálnia a kvantumfölényt. A munka során megvizsgáljuk a két terület átfedéséből származó lehetőségeket: klasszikus adatok kvantumos feldolgozását illetve a klasszikus gépi tanulás segítségével történő kvantumos hibajavítást.

Gépi tanulási algoritmusok vizsgálata kvantumszámítógépeken

Nagy Dániel[1]

[1]Eötvös Loránd Tudományegyetem, Fizika Intézet, H-1117, Pázmány Péter sétány 1/A.
Budapest, Magyarország

Elkezdve: December 1, 2019

Utolsó frissítés: 2021. március 6.

Kivonat

A modern számítástechnika jelentős eredményei közé tartozik a gépi tanulás és mesterséges intelligencia alapvető algoritmusainak kifejlesztése és ezek hasznosságának tesztelése különböző feladatokon. Ugyanakkor az elmúlt években a kvantumszámítás is jelentős fejlődésen ment keresztül, olyannyira, hogy 2019-ben a Google kísérleti csapatának sikerült demonstrálnia a kvantumfölényt. A munka során megvizsgáljuk a két terület átfedéséből származó lehetőségeket: klasszikus adatok kvantumos feldolgozását illetve a klasszikus gépi tanulás segítségével történő kvantumos hibajavítást.

Contents

List of Figures

1 Introduction

2 Quantum computing

3 Machine learning

3.1 Reinforcement learning

3.1.1 Proximal Policy optimization

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)} = \log \pi_\theta(a_t|s_t) - \log \pi_{\theta_{\text{old}}}(a_t|s_t) \tag{1}$$

$$\delta_t = r_t + \gamma V^\pi(s_{t+1}) - V^\pi(s_t) \tag{2}$$

$$\hat{A}_t = \sum_{l=0}^{T-t-1} (\gamma\lambda)^l \delta_{t+l} = (\gamma\lambda)^0 \delta_t + (\gamma\lambda)\delta_{t+1} + ... + (\gamma\lambda)^{T-t-1}\delta_{T-1} \tag{3}$$

$$L^{CLIP}(\theta) = \mathop{\mathbb{E}}_t \left[\min\left(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right] = \mathop{\mathbb{E}}_t \left[\min\left(r_t(\theta)\hat{A}_t, g(\epsilon, \hat{A}_t)\right)\right] \tag{4}$$

$$g(\varepsilon, \hat{A}_t) = \begin{cases} (1+\epsilon)\hat{A}_t, & \hat{A}_t \geq 0 \\ (1-\epsilon)\hat{A}_t, & \text{otherwise} \end{cases} \tag{5}$$

$$V^\pi_{\text{targ}}(s_t) = \sum_{l=0}^{T-t} \gamma^l r_{t+l} \tag{6}$$

$$L^{VF} = \mathop{\mathbb{E}}_t \left[\left(V^\pi(s_t) - V^\pi_{\text{targ}}(s_t)\right)^2\right] \tag{7}$$

$$H[\pi] = \mathop{\mathbb{E}}_t \left[-\sum_{a\in\mathcal{A}} \pi_\theta(a|s_t) \log \pi_\theta(a|s_t)\right] \tag{8}$$

$$L = L^{CLIP} + c_1 L^{VF} + c_2 H[\pi] \tag{9}$$

How to measure the entropy term for a quantum state ρ? How to measure the KL-divergence of two quantum states ρ_1, ρ_2?

Algorithm 1 PPO-Clip

procedure PPOCLIP(ϵ, E, N, T, K)
  for all $i \in \{1, ..., E\}$ **do**
    for all $n \in \{1, ..., N\}$ **do**
      Run the old policy π_{old} in the environment for T timesteps.
      for all $t \in \{1, ..., T\}$ **do**
        Calculate the advantage estimate $\hat{A}_t$
      end for
    end for
    for all $k \in \{1, ..., K\}$ **do**
      Sample a batch of size $M \leq NT$ and optimize the surrogate loss L.
    end for
  end for
end procedure

4 Quantum Machine Learning

4.1 Parametric quantum circuits

4.2 Calculating the gradients of the parameters

A Mathematical preliminaries

A.1 Hilbert spaces

Definition 1. ***Hilbert-space***
Given a field T (real or complex), a vector space $\mathcal{H}$ endowed with an inner product, is called a Hilbert-space, if it is a complete metric space with respect to the distance function induced by the inner product. The inner product is a map $\langle \cdot | \cdot \rangle : \mathcal{H} \times \mathcal{H} \to T$, for which $\forall x, y, z \in \mathcal{H}$:

- $\langle x|x \rangle \geq 0$
- $\langle x|x \rangle = 0 \iff x = \mathbf{0} \in \mathcal{H}$
- $\langle x|y \rangle = \langle y|x \rangle^*$, *where* * *denotes complex conjugation.*
- $\langle x|\alpha y + \beta z \rangle = \alpha \langle x|y \rangle + \beta \langle x|z \rangle$, *where* $\alpha, \beta \in T$

The norm induced by this inner product is a map $|| \cdot || : \mathcal{H} \to T$ defined as

$$||x|| = \sqrt{\langle x|x \rangle},$$

And the metric induced by this norm is defined as

$$d(x, y) = ||x - y|| = \sqrt{\langle x - y|x - y \rangle}.$$

The space $\mathcal{H}$ is said to be complete if every Cauchy-sequence is convergent with respect to the norm, and the limit is in $\mathcal{H}$. That is, each sequence $x_1, x_2, ...$, for which

$$\forall \varepsilon > 0 \; \exists N(\varepsilon) \text{ so, that } n > m > N(\varepsilon) \implies ||x_n - x_m|| < \varepsilon.$$

Definition 2. ***Linear functional***
Let $\mathcal{H}$ be a Hilbert-space over the field T. Then, the map $\varphi : \mathcal{H} \to T$ is called a linear functional, if

$$\varphi(\alpha x + \beta y) = \alpha\varphi(x) + \beta\varphi(y),\ \forall \alpha, \beta \in T,\ x, y \in \mathcal{H}.$$

Definition 3. ***Dual space***
Given a Hilbert-space $\mathcal{H}$, its dual space, $\mathcal{H}^*$ is the space of all continuous linear functionals from the space H into the base field. The norm of an element in $\mathcal{H}^*$ is

$$||\varphi||_{\mathcal{H}^*} \stackrel{def}{=} \sup_{||x||=1,\, x \in \mathcal{H}} |\varphi(x)|.$$

Theorem 1. ***Riesz representation theorem***
For every element $y \in \mathcal{H}$, there exists a unique element $\varphi_y \in \mathcal{H}^*$, defined by

$$\varphi_y(x) = \langle y|x\rangle\,,\ \forall x \in \mathcal{H}.$$

The mapping $y \mapsto \varphi_y$ is an antilinear mapping i.e. $\alpha y_1 + \beta y_2 \mapsto \alpha^*\varphi_{y_1} + \beta^*\varphi_{y_2}$, and the Riesz-representation theorem states that this mapping is an antilinear isomorphism. The inner product in $\mathcal{H}^*$ satisfies

$$\langle \varphi_x|\varphi_y\rangle = \langle x|y\rangle^* = \langle y|x\rangle\,.$$

Moreover, $||y||_{\mathcal{H}} = ||\varphi_y||_{\mathcal{H}^*}$.

Definition 4. ***Dirac-notation***
From now on, the elements in $\mathcal{H}$ will be denoted by $|x\rangle$ and their corresponding element in $\mathcal{H}^*$ as $\langle x|$.

A.2 Linear operators on Hilbert spaces

Definition 5. ***Linear operators***
A map $\hat{A} : \mathcal{H}_1 \to \mathcal{H}_2$ is a linear operator, if

$$\hat{A}(\alpha\,|x\rangle + \beta\,|y\rangle) = \alpha(\hat{A}\,|x\rangle) + \beta(\hat{A}\,|y\rangle).$$

Remark 1. *If not stated otherwise, we will assume that $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}$.*

Remark 2. *Operators will be denoted with a hat $(\hat{\cdot})$.*

Definition 6. ***Bounded linear operators***
A linear operator $\hat{A} : \mathcal{H} \to \mathcal{H}$ is bounded, if

$$\exists m \in \mathbb{R} : |\,\langle v|\hat{A}|v\rangle\,| \leq m\,\langle v|v\rangle\,,\, \forall\,|v\rangle \in \mathcal{H}$$

Remark 3. *The set of all bounded operators on $\mathcal{H}$ is denoted $\mathcal{B}(\mathcal{H})$.*

Definition 7. ***Commutators and anticommutators***
Since operators usually do not commute, its useful to define their commutator and anticommutator:

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}$$
$$\{\hat{A}, \hat{B}\} = \hat{A}\hat{B} + \hat{B}\hat{A}$$

Definition 8. ***Operator norm***
The operator norm of an operator $\hat{A}$ is defined as

$$||\hat{A}|| \stackrel{def}{=} \inf\{c \geq 0 : ||\hat{A}\,|v\rangle\,|| \leq c||\,|v\rangle\,||, \forall\,|v\rangle \in \mathcal{H}\}$$

Definition 9. ***Trace-class operators***
An operator $\hat{A}$ is called trace-class if it admits a well defined and finite trace $\mathrm{Tr}\left[\hat{A}\right] = \sum_j \langle j|\hat{A}|j\rangle$

Definition 10. ***Positive operators***
An operator $\hat{A}$ is called positive if $\langle v|\hat{A}|v\rangle \geq 0, \forall |v\rangle \in \mathcal{H}$. If $\hat{A} = \sum_j \lambda_j |j\rangle \langle j|$ then $\hat{A}$ is positive if $\lambda_j \geq 0$.

Definition 11. ***Projections*** *An operator $\Pi : \mathcal{H} \to \mathcal{H}$ is a projection if $\Pi^2 = \Pi$.*

A.3 Hermitian Operators, Unitary Operators, Spectral theorem, Hadamard-lemma

Definition 12. ***Hermitian adjoint***
Consider a **bounded** *linear operator $\hat{A} : \mathcal{H} \to \mathcal{H}$. The hermitian adjoint of $\hat{A}$ is a bounded linear operator $\hat{A}^\dagger : \mathcal{H} \to \mathcal{H}$ which satisfies*

$$\langle y| \hat{A} |x\rangle = \left(\langle x| \hat{A}^\dagger |y\rangle\right)^*, \ \forall |x\rangle, |y\rangle \in \mathcal{H}. \tag{10}$$

Definition 13. ***Hermitian operators***
A bounded linear operator $\hat{H} : \mathcal{H} \to \mathcal{H}$ is Hermitian if

$$\hat{H} = \hat{H}^\dagger, \ \textit{i.e.}\ \hat{H} |x\rangle = \hat{H}^\dagger |x\rangle, \ \forall |x\rangle \in \mathcal{H}. \tag{11}$$

Definition 14. ***Unitary operator***
A bounded linear operator $\hat{U} : \mathcal{H} \to \mathcal{H}$ is unitary if

$$\hat{U}\hat{U}^\dagger = \hat{U}^\dagger\hat{U} = 1, \ \textit{in other words,}\ \hat{U}^\dagger = \hat{U}^{-1}. \tag{12}$$

Definition 15. ***Eigenvalues and eigenvectors***
Consider bounded linear operator $\hat{A}$. If exist a vectors $|k\rangle \in \mathcal{H}$ such that

$$\hat{A} |k\rangle = \lambda_k |k\rangle, \tag{13}$$

then $|k\rangle$ is called an eigenvector of $\hat{A}$ and λk is the corresponding eigenvalue.

An important property of Hermitian operators is that they can be diagonalized with real eigenvalues. This is formally stated by the spectral theorem:

Theorem 2. ***The Spectral theorem***
Let $\hat{A}$ be a bounded Hermitian operator on some Hilbert-space $\mathcal{H}$. Then there exists an orthonormal basis in $\mathcal{H}$ which consists of the eigenvectors of $\hat{A}$ and each eigenvalue of $\hat{A}$ is real.

This means that any bounded Hermitian operator $\hat{H}$ can be decomposed as

$$\hat{H} = \sum_k \lambda_k \hat{P}_k = \sum_k \lambda_k |k\rangle\langle k| \tag{14}$$

where λ_k and $|k\rangle$ are the eigenvalues and eigenvectors of $\hat{H}$.

Definition 16. ***Exponential of operators*** *If X is a linear operator, we can define the exponential of X:*

$$e^X = \sum_{n=0}^{\infty} \frac{X^n}{n!}$$

Important: The product of exponentials of operators generally isn't equal to the exponential of their sum:

$$e^X e^Y = e^{Z(X,Y)} \neq e^{X+Y},$$

where $Z(X,Y)$ is given by the Baker-Campbell-Hausdorff formula:

$$Z(X,Y) = X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{1}{12}[Y,[X,Y]] - \frac{1}{24}[Y,[X,[X,Y]]]$$
$$- \frac{1}{720}([[[[X,Y],Y],Y],Y] + [[[Y,X],X],X],X]) + ...$$

It is however equal if $[X,Y] = 0$:

$$\text{if } [X,Y] = 0 \implies e^X e^Y = e^{X+Y}$$

There are 2 important special cases:

Theorem 3. ***The Hadamard-lemma***

$$e^X Y e^{-X} = Y + [X,Y] + \frac{1}{2!}[X,[X,Y]] + \frac{1}{3!}[X,[X,[X,Y]]] + ...$$

Theorem 4. *If X and Y commute with their commutator, i.e. $[X,[X,Y]] = [Y,[X,Y]] = 0$, then:*

$$e^X e^Y = e^{X+Y+\frac{1}{2}[X,Y]}$$

Theorem 5. *If $[X,Y] = sY$ with $s \in \mathbb{C}, s \neq 2i\pi n, n \in \mathbb{Z}$ then:*

$$e^X e^Y = \exp\left(X + \frac{s}{1-e^{-s}} Y\right)$$

A.4 Pure and mixed quantum states

Definition 17. ***Quantum states***
A quantum state of a quantum system is a mathematical entity that provides a probability distribution for the outcomes of each possible measurement on the system.

Definition 18. ***Pure quantum states***
Pure quantum states are quantum states that can be described by a vector $|\psi\rangle$ of norm 1.

If one multiplies a pure quantum state by a complex scalar $e^{i\alpha}$, then the new state is physically equivalent to the former, thus $|\psi\rangle$ and $e^{i\alpha}|\psi\rangle$ are the same pure state. The transformation $|\psi\rangle \to e^{i\alpha}|\psi\rangle$ does not change the outcomes of measurements on the state, however the phase α is important in quantum algorithms.

Example. *For example, the states $\frac{1}{\sqrt{2}}(|0\rangle + e^{i\pi}|1\rangle)$ and $\frac{1}{\sqrt{2}}(|0\rangle + e^{i\frac{\pi}{2}}|1\rangle)$ are not the same quantum state, but in both states there is 50-50 percent probability of measuring $|0\rangle$ and $|1\rangle$.*

Definition 19. ***Density Matrix***
A quantum state $\hat{\rho}$ is a trace-1, self-adjoint, positive semidefinite operator. The set of quantum states is

$$\mathcal{S}(\mathcal{H}) = \{\hat{\rho} : \hat{\rho} \geq 0, \hat{\rho} = \hat{\rho}^\dagger, \mathrm{Tr}[\hat{\rho}] = 1\}$$

A quantum state is pure if and only if $\hat{\rho}^2 = \hat{\rho}$. Also, if ρ is a pure state, then it can be written as $\hat{\rho} = |\psi\rangle\langle\psi|$. The operator ρ is called the density operator or density matrix.