

Training and Meta-Training Binary Neural Networks with Quantum Computing

Abdulah Fawaz, Paul Klein, Sebastien Piat, Simone Severini, Peter Mountney

Nicolas Melo

CIn - UFPE

nmo@cin.ufpe.br

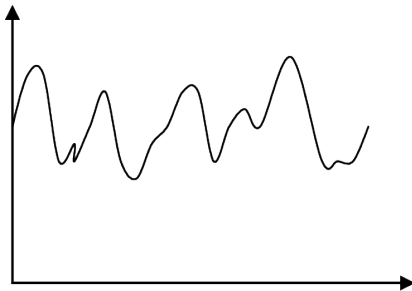
Outubro, 2019

- 1 Introdução
- 2 Trabalhos relacionados
- 3 Computação quântica
- 4 Método proposto
- 5 Resultados
- 6 Conclusões e trabalho futuro
- 7 Pontos de interesse

Introdução

Introdução

- Problema:
 - Encontrar um conjunto adequado de pesos para redes neurais
- Dificuldades:
 - Explorar espaços de busca muito grandes de forma eficiente
 - Rede pode convergir para mínimos locais
 - Escolha dos hiper-parâmetros
- Possível solução:
 - Uso de computação quântica para representar todos os possíveis parâmetros e hiper-parâmetros



- Motivação:
 - Possibilidade de representar o cenário completo para a função de custo de redes neurais como um estado quântico
- Demonstração:
 - Redes neurais binárias
 - Treinamento de uma rede por um computador quântico usando ***amplificação de amplitude***
 - Aplicar o método proposto às conectividades e pesos da rede para treinar redes neurais binárias
- Adaptação proposta:
 - Representar o cenário de meta-perda de várias arquiteturas de redes neurais simultaneamente
 - Treinar e projetar, simultaneamente, uma rede neural binária

Trabalhos relacionados

● Redes Neurais binárias

- Redes neurais com pesos e ativações assumindo apenas valores binários
- Vantagens em tempo de compressão e inferência
- Eficiência das operações bit-a-bit
- Relativamente difíceis de treinar
- Métodos de treinamento alternativos levam mais tempo do que para redes neurais não-binárias
- Ótimo desempenho em *datasets* pequenos (**MNIST** e **CIFAR10**)
- Desempenho ruim em *datasets* grandes (**ImageNet**)
 - Relaxar restrições de binarização
 - Várias ativações binárias
- Argumenta-se que uma melhor estratégia de treinamento é suficiente para alcançar uma alta acurácia
 - Alterar função de ativação
 - Diminuir taxa de aprendizado
 - Termo de regularização diferente

• Aprendizagem de Máquina Quântica

- Computação quântica:
 - Possibilidade de obter complexidade computacional significativamente melhorada em comparação com os algoritmos clássicos
 - É necessário ter uma máquina "grande" e livre de erros para manipular uma grande quantidade de dados paralelamente
- Algoritmo de Grover:
 - Algoritmo de busca em uma base de dados desestruturada
 - Amplificação de amplitude
 - Ordem $O(\sqrt{N})$, para uma base de dados de tamanho N , no lugar de $O(N)$ no caso clássico
 - Teve sua otimalidade provada
- Existem alguns algoritmos de clusterização e resolução de sistemas lineares de equações
- Maior problema é a falta de algoritmos eficientes para o treinamento de circuitos quânticos
- Proposta do artigo é diferente das já existentes para classificação binária

Computação quântica

- **Bits vs Qubits**
- Circuito quânticos
- É computacionalmente custoso simular o comportamento de *qubits* em computadores clássicos
- Algoritmos em termos de circuitos/qubits
- **Qubits**
 - $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, onde $|\alpha|^2 + |\beta|^2 = 1$
 - Conceitos relacionados: superposição, emaranhamento, medição
 - N qubits podem representar 2^N *strings* de tamanho N
- **Portas lógicas quânticas**
 - Matrizes complexas que realizam operações em qubits
 - $|\varphi\rangle = U|\psi\rangle$
 - Reversibilidade
 - Difícil analogia com a computação clássica
 - Portas lógicas usadas: X , Z , H (*Hadamard*), $CNOT$

Método proposto

Método proposto

- Superposição:
 - Representação de todos os parâmetros como qubits
- Prova de conceito:
 - Restrição a um pequeno problema
- Circuito quântico análogo às redes binárias com os parâmetros desejados armazenados nos valores dos qubits
- Propõe-se obter a mesma precisão que seu equivalente clássico
- A saída do circuito é uma superposição de todos os valores possíveis e seus respectivos resultados

$$\frac{1}{\sqrt{W}} \sum_{w \in \mathbb{W}} |w\rangle |O_w\rangle$$

- Onde \mathbb{W} é o conjunto de todos os possíveis pesos binários, W é seu tamanho e O_w é a saída da rede neural, dado o conjunto de pesos w

- Todo o cenário de custos está contido nesse estado quântico e é gerado com apenas um único circuito quântico
- Uma rede meta-neural com as conexões representadas por parâmetros binários também pode ser criada
- O estado de saída do circuito quântico conterá todo o cenário de meta-custo de todos os pesos possíveis e todas as conexões possíveis de uma rede neural simultaneamente
- Experimento:
 - Rede neural em formato de circuito
 - Pesos em superposição
 - O estado resultado é usado para treinar a rede neural

- Experimento:

- Dois problemas com 3 características binárias - 2^3 pontos:

- 1. $y(x_1, x_2, x_3) = \text{sign}(x_3x_1 + x_2)$
- 2. $y(x_1, x_2, x_3) = \text{sign}(x_1 + x_2 + x_3)$
- $\text{sign}(x) = \begin{cases} 1, & \text{se } x \geq 0 \\ -1 & \text{caso contrário} \end{cases}$

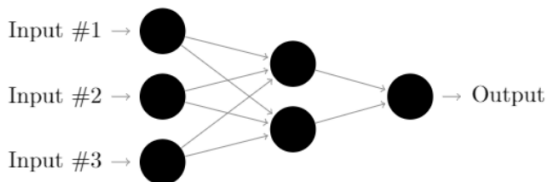


Figure 1: The structure of the BNN we are training. It has a total of eight binary weights.

- **Representando valores numéricos:**

- Pesos $(+1, -1)$ e valores binários $(+1, 0)$ são representados pelos estados $|1\rangle$ e $|0\rangle$

- **Multiplicando valores por pesos binários:**

- Para multiplicar um qubit que representa um valor binário por um qubit representando um peso binário é necessário usar uma porta ***anti-CNOT*** [CNOT ativada pelo 0], onde o peso é usado como controle

- **Implementando a função de ativação:**

- Descobrir se existem $N/2$ qubits no estado $|1\rangle$
- Função majoritária, substituindo portas AND por CCNOT (*Toffoli*) e portas OR por CNOT e NOT

- **Calculando a acurácia:**

- Executar a QBNN com os pesos em superposição para cada ponto do conjunto de treinamento separadamente
- Para um *dataset* de tamanho N , temos um registrador de N qubits contendo as previsões para cada um dos N pontos

Método proposto

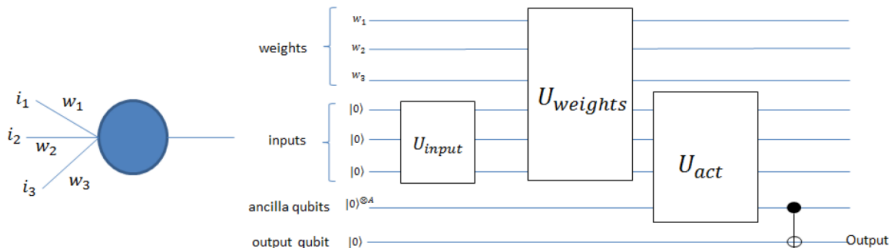


Figure 2: A quantum circuit (right) corresponding to the neuron (left). Quantum circuits are read from left to right just like classical computing circuits. The boxes represent operations on qubits. The three quantum operations mimic the classical equivalents. The data is first input, then it is weighted by the weight qubits, and finally some activation function is implemented. The final symbol on the bottom right represents a CNOT gate and stores our result in a dedicated output qubit. Ancilla qubits are additional 'helper' qubits needed to perform certain operations.

- Treinamento quântico a partir de um cenário de custo:
 - ***Amplificação de amplitude:***
 - Treinar a QBNN é como uma busca por um estado dentro da superposição
 - Amplificar a amplitude de probabilidade do estado desejado
 - Operador de amplificação Q busca aumentar a probabilidade de se obter estados que correspondam a uma alta acurácia
 - $Q \equiv -U_{QBNN} S_0 U_{QBNN}^{-1} S_X$
 - U_{QBNN} é o circuito, S_0 e S_X invertem o sinal das amplitudes no estado inicial e do estado alvo, respectivamente
 - Q precisa ser aplicado repetidamente para encontrar o estado-alvo

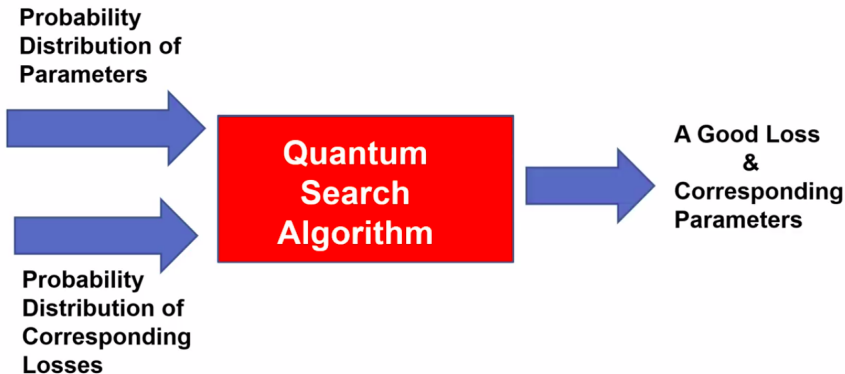
Classical Neural Network



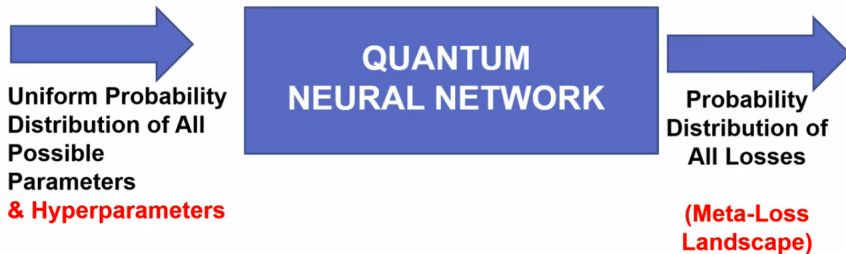
Quantum Neural Network



Training a Quantum Neural Network

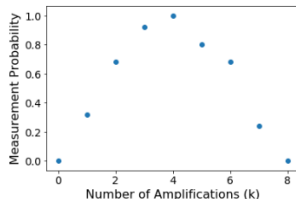


Meta Training a Quantum Neural Network

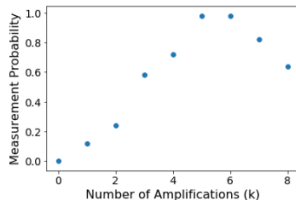


Resultados

- Os resultados foram obtidos através da simulação da QBNN no **ProjectQ** - <https://github.com/ProjectQ-Framework/ProjectQ>
- Para os dois problemas dados:
 - Probabilidade de sucesso no treinamento acima de 90% depois de 5 e 6 iterações (problema 1 e problema 2)
 - Classicamente, seriam necessárias 28 e 57 iterações, respectivamente
 - 100 execuções do algoritmo
- Treinamento com uma QBNN mais complexa, com indicador de presença/ausência de um conjunto de conexões
 - A saída é um cenário de meta-custo
 - Objetivo é encontrar um conjunto apropriado de pesos e conexões
 - Aprendizado apenas para a primeira camada, com a segunda fixa
 - Foi suficiente realizar entre 16 e 20 amplificações para obter uma probabilidade razoável

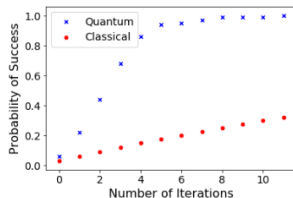


(a)

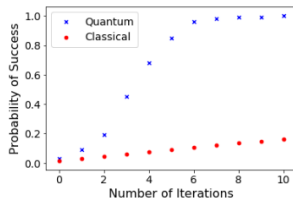


(b)

Figure 4: The relationship between the probability of obtaining an optimal set of weights against the number of quantum amplifications, k , for problem 1 (a) and problem 2 (b). Each point represents a simulation of 50 separate runs of the algorithm at the given k and the probability of success of those 50 runs.



(a) Problem 1



(b) Problem 2

Figure 5: A plot comparing the scaling of a quantum search algorithm over a classical one. The quantum data is the cumulative probability of success over 100 runs of the algorithm. Classical results are analytically derived from the known probability of obtaining a solution by random search. The superior scaling of the quantum algorithm becomes more prominent for harder problems.

Conclusões e trabalho futuro

Conclusões e trabalho futuro

- Método vantajoso, pois possui aceleração quadrática em relação ao análogo clássico
- O método apresentado não possui limitações
- É possível redefinir o operador Q para impedir o *overfitting*
- Também é possível utilizar um *batch* diferente para permitir usar menos qubits durante a execução
- Para aceitar entradas não-binárias e melhorar a escala da função de ativação é necessário melhorar/substituir esta por uma mais eficientemente implementável em um computador quântico
- Aplicar computação quântica em problemas de aprendizado de máquina ainda é uma área de pesquisa em aberto

Pontos de interesse

- Simulação do algoritmo quântico em computador clássico
 - Limitação na quantidade de qubits disponíveis
 - Limitação no número de portas que podem ser aplicadas
- O algoritmo de Grover não é muito bom para o caso geral
- Os autores ressaltam que o método proposto não se restringe a redes binárias, mas não é exposto, explicitamente, um caminho para aplicação do método em redes com pesos reais
- Apesar de os autores pontuarem que a única limitação da aplicação do método proposto é o hardware quântico atual, outras questões surgiram, pontuados no *slide 26*
 - *"While our experiment was only a small proof of concept, lacking the complexity of current data science problems, this limitation came only from limits within available resources, not the method itself. From the binarisation, to the NN architecture, to the data size, all were limited only by the inability to simulate large qubit numbers."*

Training and Meta-Training Binary Neural Networks with Quantum Computing

Abdulah Fawaz, Paul Klein, Sebastien Piat, Simone Severini, Peter Mountney

Nicolas Melo

CIn - UFPE

nmo@cin.ufpe.br

Outubro, 2019