Síntese de Unitárias em Circuitos de Clifford e Clifford+T: uma Abordagem com RL para 2 qubits

Poliana N. Ferreira, Fabrício de Souza Luiz e Marcos César de Oliveira

***Resumo*— Este trabalho investiga o uso de aprendizado por reforço para geração de circuitos quânticos equivalentes a unitárias de 1 e 2 qubits, com foco nos grupos Clifford e Clifford+T. O agente interage com o ambiente escolhendo portas, avaliando a fidelidade em relação à unitária alvo e recebendo recompensas que equilibram precisão e operações. Nos testes, modelos treinados para Clifford alcançaram 1.0 de fidelidade, enquanto Clifford+T apresentou maior dificuldade, maior tempo de convergência e fidelidades médias de 0.98, variando entre 0.83 e 1. Os resultados evidenciam o potencial e os desafios do RL na síntese quântica.**

***Palavras-Chave*— Computação Quântica, Síntese de Circuitos, Aprendizado por Reforço**

***Abstract*— This work investigates the use of reinforcement learning for the generation of quantum circuits equivalent to unitary operations of 1-2 qubits, focusing on Clifford and Clifford+T groups. The agent interacts with the environment by selecting gates, evaluating fidelity with respect to the target unitary, and receiving rewards that balance accuracy and circuit size. During tests, models trained with Clifford targets achieved perfect fidelity (1.0), while Clifford+T showed longer convergence times, and average fidelities of 0.98, ranging from 0.83 to 1. The results highlight the potential and the challenges of RL in quantum circuit synthesis.**

***Keywords*— Quantum Computing, Circuit Synthesis, Reinforcement Learning**

I. Introdução

A computação quântica vem avançando rapidamente e com ela cresce a necessidade de métodos sistemáticos para a construção de circuitos que implementem operações unitárias específicas. As operações quânticas são representadas por matrizes unitárias e, em princípio, qualquer operação pode ser decomposta em um conjunto finito de portas elementares. O conceito de universalidade de portas garante que certos conjuntos, como $H, T, CNOT$ ou $R_y, R_z, CNOT$, permitem a síntese de qualquer operação unitária como um certo grau de proximidade. No entanto, construir um único modelo capaz de gerar circuitos para todas as possíveis unitárias é uma tarefa exponencialmente difícil [3].

Na literatura, observa-se a estratégia de restringir o problema a subconjuntos estruturados de operações. Isto é, trabalhar com grupos específicos de unitárias, definidos a partir de portas bem conhecidas e com propriedades úteis [1], [2]. Esse enfoque permite testar e validar metodologias de forma mais controlada, além de revelar a complexidade associada a diferentes classes de operações.

Poliana N. Ferreira, DFMC, UNICAMP, Campinas-SP, e-mail: p224317@dac.unicamp.br; Fabrício de Souza Luiz, DFMC, UNICAMP, Campinas-SP, e-mail: fsluiz@unicamp.br; Marcos César de Oliveira, DFMC, UNICAMP, Campinas-SP, e-mail: mcoliv@unicamp.br

II. Grupo de Clifford e Clifford+T

Um dos casos mais estudados é o grupo de Clifford, formado por operações que podem ser geradas a partir das portas Hadamard (H), de fase (S) e CNOT [1]. Embora esse grupo tenha importância central em protocolos de correção de erros e estabilizadores, ele não é universal para computação quântica. Para alcançar universalidade, é necessário acrescentar a porta T, formando o conjunto Clifford+T [3], [2].

Um ponto relevante é que, para avaliar métodos de síntese, uma estratégia é não considerar unitárias totalmente aleatórias, mas sim aquelas que surgem de composições finitas desses grupos. Assim, a dificuldade do problema está diretamente ligada ao número de portas necessárias para gerar a operação desejada, refletindo a complexidade prática da decomposição.

III. Aprendizado por Reforço

O aprendizado por reforço (Reinforcement Learning, RL) é um paradigma de aprendizado de máquina no qual um agente interage com um ambiente, explorando ações e recebendo recompensas que guiam seu processo de otimização [4]. Entre os métodos modernos, destaca-se o Proximal Policy Optimization (PPO), que equilibra estabilidade de treinamento e eficiência de atualização da política. Nesta pesquisa, optou-se por utilizar o Python e a biblioteca Stable Baselines 3 para implementação.

Nesse contexto, o RL se mostra promissor para a síntese de circuitos quânticos: o agente escolhe portas a serem aplicadas, o ambiente avalia a fidelidade do circuito aproximado em relação à unitária alvo e a recompensa direciona o processo para soluções mais compactas e precisas.

IV. Métodos

No estudo inicial descrito aqui, o aprendizado por reforço foi aplicado à tarefa de geração de circuitos quânticos equivalentes a operações unitárias alvo, restringindo a análise aos casos de 1 e 2 qubits. O método foi estruturado a partir da definição dos elementos centrais do ambiente de interação, seguindo o paradigma clássico de aprendizado por reforço.

A. Base de Dados

A geração da base de dados foi realizada considerando explicitamente os grupos de portas Clifford e Clifford+T, para cada base de treino e teste, de modo a permitir a análise comparativa do desempenho do agente em diferentes cenários.

Foi introduzida variação no nível de dificuldade das instâncias, parametrizada pelo número de portas necessárias para gerar a unitária, o qual variou de 1 a 15. O número de exemplos gerados foi incremental para cada dificuldade, em

razão da limitação natural do número de combinações. Ao final, foram gerados 7000 exemplos de unitárias geradas na dificuldade 15.

B. Definições para o Aprendizado por Reforço

O agente é a entidade responsável por escolher ações com base em um estado observado, seguindo uma política que é atualizada ao longo do treinamento. Seu objetivo é aplicar uma sequência de portas quânticas que aproxime a unitária alvo, equilibrando fidelidade e eficiência (isto é, o número de portas utilizadas).

As ações correspondem à escolha de uma porta quântica a ser aplicada. No caso desse estudo, as possibilidades de ações foram as portas Clifford ou Clifford+T a depender do teste.

O ambiente corresponde ao circuito quântico em construção, representado no framework *Gymnasium*. A cada ação escolhida pelo agente, o circuito é atualizado e a operação aproximada se modifica, aproximando-se ou afastando-se da unitária alvo.

Já o estado observado pelo agente é definido pela relação entre a unitária aproximada até o momento (U_{aprox}) e a unitária alvo (U_{target}). Formalmente, a observação foi dada pela concatenação das partes real e imaginária da matriz $U_{\text{aprox}}^{\dagger} U_{\text{target}}$, achatada em um vetor:

$$\text{obs} = \begin{bmatrix} \text{Re}\left(U_{\text{aprox}}^{\dagger} U_{\text{target}}\right) \\ \text{Im}\left(U_{\text{aprox}}^{\dagger} U_{\text{target}}\right) \end{bmatrix}_{\text{flatten}}.$$

Essa escolha evita ambiguidades: se o estado fosse definido apenas por U_{aprox}, o agente poderia assumir erroneamente que a mesma porta deve sempre ser aplicada em determinada configuração intermediária, ignorando o objetivo final.

C. Métricas e Recompensa

A aproximação entre a unitária alvo U e a unitária construída V foi avaliada pela fidelidade:

$$\text{fidelidade}(U, V) = \frac{|\operatorname{Tr}(U^{\dagger} V)|^2}{d^2},$$

com $d = 2^n$. Foi adotado como limiar de satisfação o valor 0.98. Cada episódio foi limitado a 750 ações.

A recompensa foi definida como:

$$R_t = 10 \cdot (\text{fid}_t - \text{fid}_{t-1}) - 0.01 * (\text{step}_t),$$

incentivando ganhos sucessivos de fidelidade e penalizando o uso excessivo de portas. Esse esquema inicial garante progresso contínuo, embora futuras melhorias possam refinar a função de recompensa.

V. Resultados

Os experimentos envolveram o treinamento e teste de modelos baseados em aprendizado por reforço para aproximar unitárias geradas a partir de circuitos de diferentes níveis de dificuldade nos grupos Clifford e Clifford+T.

Durante o treinamento, cada episódio se inicia com a unitária identidade. O agente escolhe inicialmente ações aleatórias, aplicando portas quânticas e modificando progressivamente a unitária aproximada. A cada ação, o agente recebe recompensas que equilibram fidelidade e número de portas, até atingir o limite de 750 ações ou alcançar a fidelidade desejada. Ao final do episódio, um novo é iniciado, permitindo ao agente explorar e explotar (*exploration–exploitation trade-off*) até aprender quais sequências de ações são mais eficazes em cada etapa do processo.

O modelo gerado foi testado em instâncias nunca vistas durante o treinamento, mas construídas da mesma forma (portas Clifford ou Clifford+T, e variação de dificuldade). Dessa forma, garante-se que o teste avalia a capacidade de generalização do agente.

Nos experimentos com unitárias de Clifford de 1 qubit, todos os modelos, em todos os níveis de dificuldade (1 a 15 portas), atingiram fidelidade de 100% no conjunto de teste, reproduzindo a sequência mínima de portas necessárias. Resultados análogos foram obtidos para o caso de 2 qubits.

No entanto, ao incluir a porta T (Clifford+T), o desempenho caiu. O tempo de convergência aumentou consideravelmente, exigindo maior número de episódios para estabilização do aprendizado, e nem todos os modelos alcançaram a fidelidade ótima. Em média, os testes com 1 e 2 qubits resultaram em fidelidade de aproximadamente 0,98, mas com variação entre 0,83 e 1, dependendo do nível de dificuldade e do exemplo.

VI. Considerações Finais

A diferença de desempenho entre os grupos pode ser explicada pela natureza intrínseca de cada conjunto de portas. O grupo de Clifford não é universal e possui estrutura algébrica bem definida e pode ser eficientemente simulado em computadores clássicos. Isso o torna mais previsível para o agente de aprendizado por reforço, que encontra padrões na aplicação de portas. Já no caso de Clifford+T, há aumento do espaço de busca e a complexidade do problema, mesmo para dificuldades não muito grandes.

Essa observação reflete a dificuldade do problema geral: aproximar circuitos quânticos equivalentes a qualquer unitária arbitrária. O desafio real envolve unitárias geradas por diferentes combinações de portas, em níveis variados de complexidade e em dimensões crescentes.

O presente trabalho explorou os casos iniciais de 1 e 2 qubits, mas aponta para a necessidade de investigar tanto conjuntos universais (como Clifford+T), outros grupos de interesse e unitárias quaisquer. Essa linha de pesquisa é útil para avaliar o potencial do aprendizado por reforço como ferramenta geral de síntese de circuitos quânticos e está sendo desenvolvida mais aprofundadamente pelo grupo.

Referências

[1] D. Kremer, V. Villar, H. Paik, I. Duran, I. Faro e J. Cruz-Benito, "Practical and efficient quantum circuit synthesis and transpiling with reinforcement learning,"*arXiv preprint arXiv:2405.13196*, 2024.

[2] M. Kölle, T. Schubert, P. Altmann, M. Zorn, J. Stein e C. Linnhoff-Popien, "A reinforcement learning environment for directed quantum circuit synthesis,"*arXiv preprint arXiv:2401.07054*, 2024.

[3] M. A. Nielsen e I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[4] R. S. Sutton, *Reinforcement Learning: An Introduction*. A Bradford Book, 2018.