

Transformer 기반 강화학습을 이용한 가변 노드 수 양자 네트워크에서의 얹힘 분배 최적화

석영준¹, 강대건², Nurlanbek kyzzy Asel¹, 한연희^{1†}

¹한국기술교육대학교 컴퓨터공학과 미래융합공학전공

²한국기술교육대학교 컴퓨터공학과 컴퓨터공학전공

{dsb04163, ahicraft1937, chil1207, yhhan}@koreatech.ac.kr

Transformer Based Reinforcement Learning for Entanglement Distribution Optimization in Quantum Networks with Variable Numbers of Nodes

Yeong-Jun Seok¹, Dae-Gun Kang², Nurlanbek kyzzy Asel¹, Youn-Hee Han^{1†},

¹Future Convergence Engineering, Dept. of Computer Science and Engineering, KOREATECH

²Dept. of Computer Science and Engineering, KOREATECH

요약

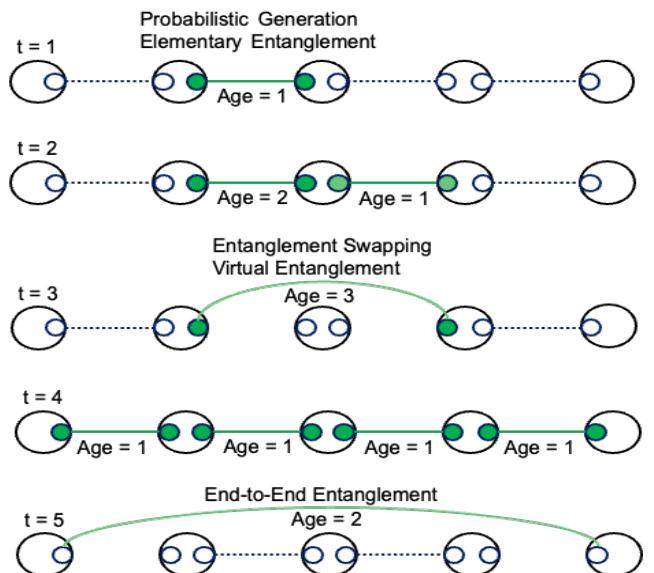
양자 네트워크에서 얹힘 분배는 원거리 양자 통신과 분산 양자 컴퓨팅을 위한 핵심 문제이며, 확률적인 얹힘 생성과 제한된 양자 자원으로 인해 복잡한 의사결정 문제로 나타난다. 기존 강화학습 기반 얹힘 분배 연구들은 고정된 노드 수 환경에서 학습이 이루어져, 네트워크 크기가 변화할 경우 정책을 그대로 적용하기 어렵다는 한계를 가진다. 이에 본 논문에서는 노드 수 변화에 강건하게 동작할 수 있는 Transformer 기반 강화학습 알고리즘을 제안한다. 제안하는 방법은 노드 수에 독립적인 상태 표현과 Transformer 기반 정책 구조를 통해 단일 학습으로 다양한 크기의 양자 네트워크에 적용 가능하다. 실험 결과, 제안된 모델은 학습에 포함되지 않은 노드 수 범위 및 구조적 패턴을 가지는 네트워크 환경에서도 재학습 없이 안정적인 얹힘 분배 성능을 보였다.

I. 서론

양자 네트워크는 원거리 양자 통신과 분산 양자 컴퓨팅을 가능하게 하는 핵심 인프라로 주목받고 있으며[1], 이에서 얹힘 분배 (entanglement distribution)는 네트워크 성능과 신뢰도를 좌우하는 핵심 문제이다. 얹힘 생성과 얹힘 교환(entanglement swapping)은 확률적으로 이루어지고, 제한된 양자 메모리와 decoherence로 인해 얹힘 분배는 복잡한 의사결정 문제로 나타난다.

기존 연구들은 주로 rule-based 방식에 기반한 얹힘 분배 전략을 사용해 왔으나, 이러한 접근법은 네트워크 구조나 환경 변화에 대한 적응성이 낮다는 한계를 가진다[2]. 이를 보완하기 위해 최근에는 강화학습(Reinforcement Learning)을 활용하여 얹힘 분배 전략을 자동으로 학습하는 연구들이 제안되었으며[3], 기존 rule-based 방법 대비 우수한 성능을 보인 바 있다. 그러나 대부분의 강화학습 기반 접근법은 고정된 노드 수 환경에서 학습이 이루어져, 네트워크 크기가 변화할 경우 정책을 그대로 적용하기 어렵다는 문제가 있다.

본 연구에서는 이러한 한계를 해결하기 위해 Transformer 기반 강화학습 정책을 제안한다. Transformer 모델은 self-attention 메커니즘을 통해 입력의 개수와 순서에 독립적인 표현이 가능하므로[4], 노드 수가 서로 다른 네트워크 환경에서도 동일한 정책을 적용할 수 있다. 본 논문에서는 얹힘 분배 문제를 강화학습 문제로 정식화하고, 노드 정보를 Transformer 기반 구조로 표현함으로써 단일 학습으로 다

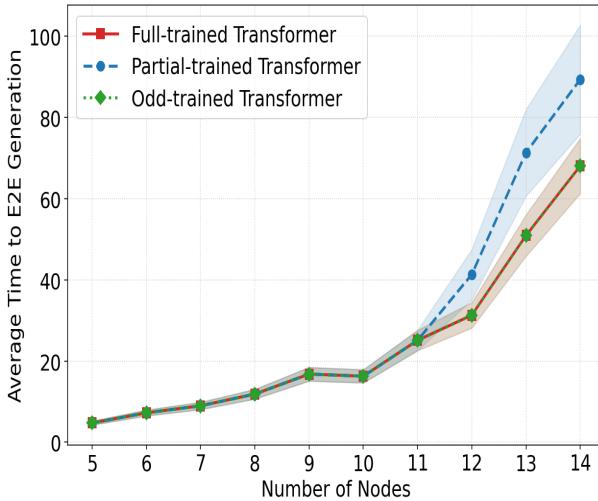


[그림 1] 선형 양자 네트워크에서 얹힘 생성과 교환 예시
양한 노드 수의 양자 네트워크에 적용 가능한 모델을 설계한다.

실험 결과, 제안된 Transformer 기반 강화학습 모델은 학습에 포함되지 않은 노드 수 환경에서도 재학습 없이 안정적인 얹힘 분배 성능을 보였으며, 이를 통해 노드 수 변화에 대한 일반화 가능성을 실험적으로 확인하였다.

II. 문제 정의

[†]한연희(Youn-Hee Han, yhhan@koreatech.ac.kr): 교신저자



[그림 2] 노드 수에 따른 E2E 생성까지의 평균 소요 시간

본 연구에서는 선형 형태로 연결된 양자 네트워크에서의 얹힘 분배 문제를 고려한다. 네트워크는 여러 노드로 구성되며, 인접 노드 간에서는 확률적으로 elementary entanglement가 생성된다. 생성된 얹힘은 제한된 양자 메모리에 저장되며, decoherence 및 cut-off 규칙에 의해 소멸될 수 있다. 중간 노드에서는 양쪽에 얹힘이 존재할 경우 얹힘 교환 (entanglement swapping)을 수행하여 원거리 노드 간 얹힘을 생성할 수 있다.

이러한 물리적 제약 하에서 얹힘 분배는 시간에 따라 변화하는 동적 의사결정 문제로 나타난다. 목표는 네트워크의 양 끝단 노드 사이에 end-to-end entanglement를 가능한 한 빠르게 생성하는 것으로, 각 시간 단계마다 중간 노드에서 얹힘 교환을 수행할지 여부를 결정하는 분배 정책을 적절히 선택해야 한다. 이 선택은 전체 분배 시간과 성공 확률에 직접적인 영향을 미친다.

기존 얹힘 분배 연구들은 대부분 특정 노드 수와 고정된 네트워크 구조를 가정하여 분배 전략을 설계해 왔다. 그러나 실제 양자 네트워크 환경에서는 노드 수 변화가 빈번하므로, 노드 수에 독립적으로 적용 가능한 일반화된 분배 전략이 요구된다. 이러한 관점에서 본 연구는 노드 수 변화에 강건한 얹힘 분배 최적화 방법을 목표로 하며, 다음 절에서는 이를 구현하기 위한 알고리즘을 제안한다.

III. 제안하는 알고리즘

본 연구에서는 얹힘 분배 문제를 해결하기 위해 Transformer 기반 강화학습 알고리즘을 제안한다. 상태는 각 노드가 현재 보유한 얹힘의 경과 시간으로 구성되며, 이를 통해 네트워크 전반의 얹힘 분포와 decoherence 영향을 압축적으로 표현한다. 행동은 각 시간 단계에서 중간 노드가 얹힘 교환을 수행할지 여부를 결정하는 이진 선택으로 정의되며, 여러 중간 노드에 대해 동시에 선택이 가능하다. 보상은 end-to-end entanglement가 생성될 경우 양의 값을 부여하고, 그 외에는 시간 지연을 반영한 작은 음의 값을 부여함으로써 전체 분배 시간을 최소화하도록 유도한다.

상태 전이는 얹힘 분배의 물리적 과정을 반영하여 진행된다. 선택된 중간 노드에서 얹힘 교환이 수행되면 기존 얹힘은 제거되고 새로운 얹힘이 생성되며, 이후 decoherence 및 cut-off 규칙에 따라 오래된 얹힘이 소멸되고 남은 얹힘의 경과 시간이 증가한다. 마지막으로 인접 노드

쌍에 대해 elementary entanglement 생성이 확률적으로 시도되며, 이러한 전이 구조는 네트워크 크기에 관계없이 동일하게 유지된다.

정책은 노드 상태를 토큰으로 입력받는 Transformer 모델을 통해 학습되며, self-attention 메커니즘을 통해 노드 간 관계를 고려한 의사 결정을 수행한다. 이를 통해 학습에 포함되지 않은 노드 수 환경에서도 동일한 정책을 적용할 수 있으며, 정책 학습에는 Proximal Policy Optimization(PPO)을 사용하였다[5].

IV. 실험

본 절에서는 제안한 Transformer 기반 강화학습 모델의 노드 수 일반화 성능을 평가한다. 성능 지표로는 end-to-end entanglement 생성 까지의 평균 시간을 사용하였으며, 모든 결과는 각 환경에 대해 100회의 독립적인 테스트 평균값으로 산출되었다. 그림 2는 서로 다른 학습 범위를 갖는 Transformer 기반 모델들의 성능을 노드 수에 따라 비교한 결과를 보여준다. Partial-trained 모델은 제한된 노드 수 환경에서만 학습되었음에도 불구하고, 학습에 포함되지 않은 노드 수 환경에서도 안정적인 성능을 보였다. 또한 홀수 노드 환경만을 학습한 모델 역시 전체 노드를 학습한 모델과 유사한 성능을 나타냈다. 이는 제안된 모델이 노드 수 변화와 패턴에 대해 재학습 없이 일반화될 수 있음을 보여준다.

V. 결론

본 논문에서는 노드 수 변화에 강건한 Transformer 기반 강화학습 얹힘 분배 알고리즘을 제안하였다. 제안된 방법은 단일 학습으로 다양한 네트워크 크기에 적용 가능하며, 실험을 통해 학습에 포함되지 않은 노드 수 환경에서도 안정적인 성능을 보임을 확인하였다. 이는 네트워크 규모가 동적으로 변화하는 실제 양자 네트워크 환경에서의 활용 가능성을 시사한다.

ACKNOWLEDGMENT

이 논문은 2023년도 및 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2023R1A2C1003143 & No. NRF-2018R1A6A1A03025526).

참고문헌

- [1] 오현수, 윤승호, 정유신, 정혜단, 허준. (2023). 양자인터넷 기본 구조 및 개요. *한국통신학회지(정보와통신)*, 40(8), 3-11.
- [2] Iñesta, Á.G., Vardoyan, G., Scavuzzo, L. et al. Optimal entanglement distribution policies in homogeneous repeater chains with cutoffs. *npj Quantum Inf* 9, 46 (2023). <https://doi.org/10.1038/s41534-023-00713-9>
- [3] Haldar, S., Barge, P. J., Khatri, S., Lee, H. (2024). Fast and reliable entanglement distribution with quantum repeaters: Principles for improving protocols using reinforcement learning. *Physical Review Applied*, 21(2), 024041.
- [4] Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).
- [5] Schulman, John, et al. "Proximal policy optimization algorithms." *arXiv preprint arXiv:1707.06347* (2017).