대규모 산업망의 QoS 보장을 위한 CVaR 기반 위험 민감 강화학습 라우팅 기법

장여진, 백민우, 길기훈, 이서영, 유지오, 이상금*
*국립한밭대학교

{jyeoj251, bmw5779, minegihun, syoung2353, uzo7383}@gmail.com, *sangkeum@hanbat.ac.kr

CVaR-Based Risk-Sensitive Reinforcement Learning Routing for QoS Assurance in Large-Scale Industrial Networks

Yeojin Jang, Minu Baek, Gihun Gil, Seoyoung Lee, Jio Yoo, Sangkeum Lee*
*Hanbat National University

요 약

산업용 연속 공정 시스템에서는 네트워크 혼잡으로 인한 꼬리 지연(Tail Latency)이 공정 품질 저하 및 설비 손상과 같은 치명적인 결과를 초래할 수 있다. 이에 본 논문은 실제 제지공정 로그 데이터 기반의 트래픽 환경에서 데이터 전송의 안정성을 보장하기 위한 위험 민감(Risk-Sensitive) 강화학습 라우팅 기법을 제안한다. 제안하는 기법은 과거 네트워크 상태를 누적한 관측을 통해 병목의 사전 징후를 포착하며 CVaR(Conditional Value-at-Risk) 기반의 PPO 학습을 적용하여 최악의 혼잡 상황에서도 성능 저하를 최소화하도록 설계하였다. 다양한 네트워크 규모와 동적 병목 시나리오에서의 실험 결과, 제안 기법은 정상 구간에서는 기존 링크 부하 기반 동적 라우팅과 동등한 효율을 유지하면서 병목 구간에서는 평균 지연과 QoS 위반율을 효과적으로 억제함을 확인하였다. 이를 통해 불확실성이 높은 산업 네트워크 환경에서 제안 기법이 높은 강건성과 신뢰성을 제공함을 입증하였다.

Ⅰ. 서 론

제지공정과 같은 연속 공정 시스템은 수많은 센서 데이터와 제어 명령이 동시에 발생하며 공정 단계별 및 트래픽 유형에 따라 지연 요구사항과 신뢰성 수준이 다르다. 이러한 환경에서 네트워크 혼잡으로 인해 발생하는 꼬리 지연(Tail Latency)은 품질 저하, 설비 손상과 같은 치명적인 결과로 이어질 수 있다[1]. 따라서 예기치 못한 혼잡·병목 상황에서 지연 및 QoS(Quality of Service) 위반을 억제할 수 있는 라우팅 기법이 요구된다.

본 논문은 제지공정 로그 데이터를 기반으로 산업망 트래픽을 구성하고 네트워크 상태를 활용하는 강화학습 기반 라우팅 기법을 제안한다. 안정성을 최우선으로 고려하여, CVaR(Conditional Value-at-Risk) 기반의 위험 민감(Risk-Sensitive) 학습 전략을 통해 최악 구간의 성능 저하를 최소화하도록 설계하였다. 제안 기법을 통해 네트워크 규모 확장 및 예기치 못한 혼잡 상황에서도 지연 및 QoS 위반의 완화를 목표로 한다.

Ⅱ. 본론

2.1 네트워크 및 트래픽 모델링

제지공정의 산업적 특성을 반영하기 위해 실제 수집된 RTDB/MES 로그 데이터를 기반으로 패킷 트래픽을 구성하였으며 태그 정보를 활용하여 공정 단계 및 트래픽 유형을 구분하고 공정별로 서로 다른 트래픽 패턴이 나타나도록 하였다.

네트워크는 패킷이 필드에서 중앙 서버로 전달되는 계층형 구조로 모델링하였다. 네트워크는 노드 집합 V와 링크 집합 E로 구성된 방향성 그래프 $G = (V, E)$로 정의하며 각 링크 $e \in E$에는 기저 지연 d_e와 용량에 해당하는 파라미터가 부여된다.

패킷 k의 종단 지연 D_k는 선택된 경로 π_k를 구성하는 링크 지연의 합으로 나타낸다.

$$D_k = \sum_{e \in \pi_k} (d_e + \omega_e)$$

여기서 ω_e는 해당 링크의 순간 부하 및 대기열 증가에 의해 추가로 발생하는 지연 성분을 의미한다. 이러한 모델은 트래픽 집중 시 평균 지연뿐 아니라 꼬리 지연이 확대되는 현상을 재현할 수 있도록 설계하였다.

2.2 강화학습 기반 라우팅 기법

패킷 전달 시점 t에서 현재 노드에 도착한 패킷은 중앙 코어 서버를 목적지로 하고 인접 노드 중 하나를 선택하여 전달된다. 이 과정은 네트워크 상태에 따라 홉을 선택하는 순차적 의사결정 문제(MDP)로 정의하며 정책은 상태를 바탕으로 다음 홉을 선택한다. 각 노드의 이웃은 기저 지연을 기준으로 정렬되며 정책은 이 중 최대 K개의 후보에 대해 이산적으로 행동을 선택한다.

정책이 혼잡·병목 상황을 인지하고 우회 결정을 내리기 위해 상태는 목적지까지의 상대적 거리 정보, 링크 부하 수준, 진행 상태로 구성된다. 혼잡은 즉시 발생하지 않고 일정 시간 누적 후 발생하는 경우가 많으므로 최근 시점의 관측을 누적한 과거 상태를 사용하여 병목의 사전 징후와 전이 패턴을 학습하도록 설계하였다.

보상은 패킷 전달 과정에서 발생하는 지연을 최소화하면서 데드라인 초과와 같이 QoS 위반을 강하게 억제하는 방향으로 정의된다. 한 스텝에서 발생한 링크 지연을 Δt라 할 때 기본 보상은 다음과 같다.

$$r_t = -\lambda_d \Delta t - \lambda_{loop} \cdot \mathbb{I}[loop]$$

여기에 누적 지연이 패킷 데드라인에 근접하거나 초과할 경우 추가 패널티를 부여하여 평균 지연뿐 아니라 지연 분포의 꼬리 구간을 직접적으로 감소시키도록 유도한다.

이러한 보상 구조를 바탕으로 정책 학습에는 PPO(Proximal Policy Optimization)를 사용한다[2]. 본 연구에서는 학습의 안정성을 보장하는 PPO를 기반으로 위험 민감 학습과 결합한다. 한 롤아웃에서 얻은 반환값 집합 중 하위 α 비율(최악 구간)에 해당하는 샘플로 정책을 업데이트하며 평균 성능보다 최악 구간 성능에 비중을 두는 학습을 수행한다. 이는 CVaR를 근사한 형태이며 다음과 같다[3].

$$CVaR_\alpha(R) = \mathbb{E}[R \mid \mathrm{R} \le Quantile_\alpha(R)]$$

이를 통해 정책은 복합적으로 발생하는 병목 상황에서 지연 폭증을 억제하는 방향으로 학습된다.

2.3 실험 및 결과 분석

제안한 위험 민감 강화학습 기반 라우팅 기법의 성능을 검증하기 위해 링크 부하 기반 동적 라우팅 기법을 비교 대상으로 설정하여 네트워크 규모 및 혼잡 시나리오 변화에 따른 지연 및 QoS 특성을 관찰한다.

파라미터	값
노드 수	50, 100, 200
병목 모델	Markov Chain
학습/평가 데이터	9,000 / 3,000 packets
업데이트 횟수	1,400
학습률 (α)	2×10^{-4}
할인율 (γ)	0.98
배치 크기	256
CVaR_α	0.25

표1. 시뮬레이션 파라미터 설정값

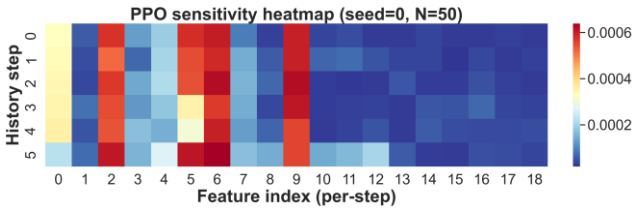

그림 1. PPO 에이전트의 민감도 히트맵

실험은 네트워크 규모 확장과 동적 병목 모델링을 통해 제안 기법의 확장성 및 환경 적응력을 검증하였다. 데이터 분리와 다중 시드 실험으로 일반화 성능을 확보했으며 특히 CVaR 기반 PPO 학습을 적용하여 최악의 혼잡 상황에서도 안정적인 성능을 유도하였다.

그림 1의 민감도 히트맵은 시계열적 관측 정보가 단순 링크 부하를 넘어 혼잡 전이 및 병목 징후를 포착하는 데 핵심적인 역할을 수행함을 보여준다. 이는 제안된 관측 설계가 에이전트에게 정상과 비정상 상황을 효과적으로 구분하여 적응적 경로 판단을 내리도록 유도했음을 의미한다.

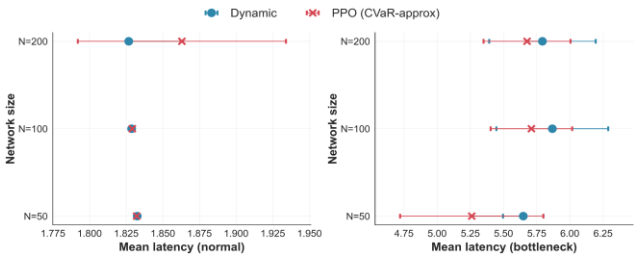

그림 2. 정상 및 병목 구간 평균 지연 비교

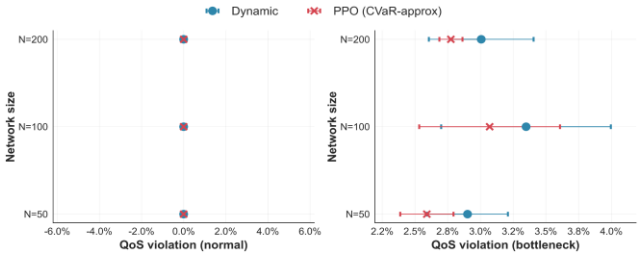

그림 3. 정상 및 병목 구간 평균 QoS 위반율 비교

실험 결과를 정상 구간과 병목 구간을 분리하여 지연 및 QoS 위반율을 비교하였다. 그림 2에서 제안 기법은 정상 구간에서 동적 라우팅과 거의 유사한 성능을 보인 반면 병목 구간에서는 PPO가 더 낮은 평균을 보이며 성능 저하 폭을 완화하는 경향이 나타난다.

그림 3은 동일한 비교를 QoS 위반율 관점에서 나타낸 것이다. 정상 구간에서는 두 방법 모두 위반율이 거의 발생하지 않지만, 병목 구간에서는 PPO가 위반율 증가를 더 억제하는 경향을 보인다. 이는 CVaR 근사 위험 민감 학습이 혼잡 환경에서 QoS 안정성을 향상시키는 데 기여한 것을 의미한다.

Ⅲ. 결론

본 논문에서는 제지공정 로그 데이터 기반의 트래픽 환경에서 네트워크 병목 시 QoS 안정성을 확보하기 위한 위험 민감 강화학습 라우팅 기법을 제안하였다. 제안 기법은 과거 상태 관측과 CVaR 기반 PPO 학습을 통해 혼잡을 사전에 포착하고 최악 구간의 성능 저하를 방지하도록 설계되었다. 실험 결과, 정상 구간에서는 기본 효율을 유지하면서 병목 구간에서는 대조군 대비 평균 지연과 QoS 위반율을 효과적으로 억제함을 확인하였다. 이를 통해 불확실성이 높은 산업용 네트워크에서 안정적인 데이터 전송을 보장하는 데 기여한다.

참 고 문 헌

[1] S. Lee *et al.*, "Anomaly detection of smart metering system for power management with battery storage system/electric vehicle," ETRI Journal, vol. 45, no. 4, pp. 650-665, 2023.

[2] K. Arulkumaran *et al.*, "Deep reinforcement learning: A brief survey," IEEE Signal Processing Magazine, vol. 34, no. 6, pp. 26-38, 2017.

[3] R. T. Rockafellar and S. Uryasev, “Optimization of conditional value-at-risk,” Journal of Risk, vol. 2, pp. 21-42, 2000.