

딜레이-민감 비디오 스트리밍 서비스를 위한 DQN 알고리즘 연구

김유노, 김현수, 오승훈, 최민석

경희대학교

rladbsh456@naver.com, nan111nan78@gmail.com, toforwitho@khu.ac.kr, choims@khu.ac.kr

A Study on Deep Q-Network Algorithm for Delay-Sensitive Video Streaming Service

Yunoh Kim, Hyeonsoo Kim, Seunghoon Oh, Minseok Choi

Kyung Hee Univ.

요약

본 논문에서는 Deep Q Network (DQN) 알고리즘을 이용하여 딜레이에 민감한 동적 비디오 스트리밍에서의 비디오 조각 전송을 조절한다. 동적 스트리밍에서는 사용자가 비디오 스트림을 chunk 단위로 수신하고, 비디오의 앞부분을 재생하면서 동시에 뒷부분의 chunk를 수신할 수 있다. 수신한 비디오 chunk는 사용자 큐에서 재생되기를 기다리고, 큐가 비게 되면 재생 지연이 발생하게 된다. 이때, 랜덤한 채널 정보를 모른 채 큐의 길이를 적당히 높여 재생 지연을 줄일 수 있도록 강화학습이 필요하며, 동시에 평균 비트율을 증가시킬 수 있는 DQN 기반 동적 스트리밍 기술을 제안한다.

I. 서론

온라인 비디오 서비스는 전체 데이터 트래픽 중 가장 많은 비율을 차지하고, 그중 동적 비디오 스트리밍 사용자 수는 급증하고 있다 [1]. 스트리밍 사용자에게 가장 중요한 성능 지표는 재생 끊김 방지와 비트율로, 네트워크 상황에 따라 동적으로 비트율을 조절하고, 이와 동시에 스케줄링 또는 자원 할당 등을 최적화하는 기술이 연구되었다.

비디오 스트림이 짧은 재생 시간을 담당하는 chunk 단위로 구성되었다는 점에 기인하여, 비디오 chunk의 전송과 비트율 조절을 함께 최적화하는 연구가 된 바 있다 [2]. 그러나 해당 연구는 채널 상황을 모두 알 때 이뤄졌다는 제한점이 있다. 이후 랜덤한 채널 상태를 모를 때 적용 가능한 강화학습 기반의 비디오 전송 [3] 또는 비트율 조절 [4] 기술이 등장했지만, 모두 자원 소모 또는 전송 지연 시간 감소에 초점을 맞추고 큐잉 시스템에서 발생하는 실제 재생 끊김 현상은 고려하지 않았다. 본 논문에서는 DQN 알고리즘을 이용하여 채널 상태를 모르더라도 재생 끊김 현상을 방지하면서 최대한 평균 비트율을 높일 수 있는 스트리밍 기술을 제안한다.

II. 딜레이-민감 비디오 스트리밍 시스템

본 논문에서는 scalable video coding (SVC)로 인코딩된 비디오 스트리밍 시스템을 고려한다. 비디오 파일은 짧은 재생 시간을 담당하는 N 개의 chunk들로 구성된다. 또한, SVC로 인코딩된 비디오의 각 chunk는 L 개의 레이어로 구성되고, 이 중 하나는 basement layer (BL)로 재생에 필수적인 레이어이며, 나머지는 모두 enhancement layer (EL)로 여러 개가 있을수록 높은 비트율을 제공할 수 있다. EL은 없더라도 재생이 가능하다. k 개의 레이어를 보낼 때의 비트율은 $r(k)$, 파일 크기는 $S(k)$ 라 하자.

송신기는 N 개의 chunk를 보낼 때 몇 개의 EL을 보낼지 선택하는 것으로 비트율을 조절할 수 있다. 시간이 $t \in \{0, 1, \dots\}$ 처럼 이산적으로 슬롯화 되었을 때, 시간 t 에서 송신기가 보내는 chunk의 수는 $M(t)$, 레이어의 수는 $K(t)$ 라 하자. 같은 시간에 보내는 모든 chunk의 레이어 수는 동일하다고 가정한다. 송신기가 채널은 Rayleigh 페이딩 채널을 가정하여, 채널 계인은 $h(t) = \sqrt{\frac{w(t)}{d^\beta}} u(t)$ 이고, $u(t) \sim CN(0, 1)$ 는 빠른 페이딩 계인, $w(t) \sim N(0, 1)$ 는 색도잉 효과, d 는 송수신기 거리, β 는 경로

손실 계수이다. 이때, 채널 용량에 따라 전송할 수 있는 비디오 chunk 수와 레이어 수의 조합이 아래와 같이 제한되며, t_c 는 chunk 재생 시간, B 는 대역폭, Γ_T 는 전송 신호대잡음비 (signal-to-noise ratio, SNR)이다.

$$M(t) \cdot K(t) \leq t_c C = B \log_2(1 + \Gamma_T |h(t)|^2) \quad (1)$$

전송된 $M(t)$ 개의 chunk들은 사용자 큐에 쌓여 재생되기를 기다린다. 시간 t 에서의 큐 길이를 $Q(t)$ 라 할 때, 시간 슬롯 길이가 chunk의 재생 시간과 같다면, 매 슬롯 1개의 chunk가 빠져나가고 전송된 $M(t)$ 개의 chunk들이 채워져서 아래와 같은 관계식을 만족한다.

$$Q(t+1) = \max\{Q(t) - 1, 0\} + M(t) \quad (2)$$

이때, $Q(t) = 0$ 이라면 재생할 chunk가 없으므로 재생 지연 또는 끊김 현상이 발생한다. 따라서, 재생 지연이 최대한 발생하지 않도록 큐 길이를 늘리려면 많은 chunk를 보내야 하지만, 이는 식 (1)에 의해 비트율을 낮춰야 하여 비디오의 평균 비트율과 재생 지연 사이에 트레이드오프가 발생한다. 이러한 트레이드오프를 조절하는 DQN 기반 스트리밍을 제안한다.

III. DQN 기반 동적 스트리밍 기술

DQN 알고리즘이란, 큐러닝에서 사용하는 오프폴리시의 장점을 가져와 타겟 네트워크를 효율적으로 학습할 수 있고, 행동 가치함수를 딥러닝을 이용해 근사하기 때문에 모든 상태와 행동에 대한 경우의 수만큼 메모리를 사용하지 않고도 더 효율적인 가치함수 도출이 가능한 알고리즘이다. 본 연구에서 고려하는 비디오 chunk 단위의 전송은 많은 상태와 행동 공간을 요구하여, DQN 알고리즘으로 학습을 진행하였다.

DQN 알고리즘을 위한 Markov Decision Process(MDP)의 상태 공간은 $Q(t)$ 와 $n(t)$ 로 구성되며, $n(t)$ 는 송신기가 시간 t 까지 몇 개의 chunk를 보냈는지를 의미한다. $Q(t)$ 를 파악하여 재생 지연을 막기 위해 더 낮은 비트율로 더 많은 chunk를 송신할지, 더 높은 비트율로 더 적은 chunk를 송신할지 결정하게 된다. 또한, $n(t)$ 값에 따라 재생 초기 또는 말기에 따라 전송 방식을 조절할 수 있게 했다.

에이전트는 해당 상태를 입력으로 받아서 취할 수 있는 행동으로 $M(t)$, $K(t)$ 를 선택한다. 시간 t 에 $M(t)$ 개의 chunk를 $K(t)$ 개의 레이어로 보내는 것이다. 단, 본 연구에서는 채널 용량을 모르는 상태에서 행동을 선택하기 때문에 행동 선택에 $C(t)$ 는 제한사항으로 적용하지 않는다.

에이전트는 시간 t 에 어떤 상태에서 어떤 행동을 선택했는지에 따라서 보상 $R(t)$ 를 받고 누적된 보상을 최대화하도록 학습한다. 보상함수는 전송하는 chunk의 비트율, 버퍼링 발생, 전송 손실에 관련된 세 가지 항목의 합으로 구성되며, $R(t) = R_1(t) + R_2(t) + R_3(t)$ 로 표현한다. 비트율에 관련된 보상은 식 (3)와 같으며, 높은 비트율로 chunk를 많이 보내면 좋은 결과를 얻을 수 있도록 $M(t)$ 과 $r(K(t))$ 를 모두 사용하였다.

$$R_1(t) = k_Q \cdot \sqrt{M(t)} \cdot r(K(t)) \quad (3)$$

버퍼링에 대한 보상은 식 (4)이며, 사용자의 큐가 비어 재생 끊김이 발생하면 큰 페널티를 주고, 큐를 일정 수준 이상 채워놓으면 재생 끊김은 방지할 수 있으므로 $Q(t)$ 값이 일정 값 이상이 되면 음의 보상이 급격히 줄어들 수 있도록 하였다.

$$R_2(t) = -k_P \cdot I[Q(t) = 0] + \min\left\{Q(t) - k_P, -\frac{k_P}{Q(t)}\right\} \cdot I[Q(t) \neq 0] \quad (4)$$

또한, 랜덤한 채널 용량을 모른 채 chunk를 전송하므로 일부 chunk는 전송에 실패하여 손실될 수 있다. 이 손실을 반영하기 위한 보상은 식 (5)이며, 채널 용량을 넘어서는 chunk의 크기에 비례한다.

$$R_3(t) = -k_L \cdot \max\{M(t) \cdot r(K(t)) - t_c C(t), 0\} \quad (5)$$

식 (3), (4), (5)의 k_Q , k_P , k_L 는 비례상수로, 각 보상 값들의 크기를 맞춰 주기 위한 용도인데, 버퍼링에 스트리밍 사용자가 가장 민감하므로 k_P 값을 다른 상수들보다 비교적 큰 값을 설정할 수 있게 한다.

보상함수 $R(t)$ 을 통해 에이전트는 채널 용량을 모르는 상태에서도 수신기의 큐가 비지 않게 하여 재생 끊김을 방지하고, 비트율을 높이면서, 송신 손실이 덜 발생하도록 송신 chunk의 개수와 비트율을 결정하게 된다. 본 연구는 DQN 알고리즘으로 정의한 MDP를 풀어 학습을 진행하였다.

IV. 실험 결과

시뮬레이션을 위해 총 299개의 chunk로 구성되고 하나의 chunk가 $t_c = 2$ sec의 재생 시간을 담당하는 Big Buck Bunny 비디오가 4개의 레이어로 인코딩될 수 있다고 가정했다 [5]. 레이어 개수 별 비트율은 0.6, 0.99, 1.5, 2.075 Mbps라 하고, $B = 5\text{MHz}$, $d = 50\text{m}$, $\beta = 2.5$ 를 사용하였다. 비교 기술로는 큐 길이에 따라 비트율을 고정적으로 결정한 상태에서 채널 용량의 평균값을 기준으로 최대한 chunk를 많이 전송하는 임계 전송 방식을 채택하였다. 이를 통해 얻어진 Fig. 1-3은 각각 SNR에 따른 재생 끊김 발생 횟수와 평균 비트율, 전송 손실량을 보여준다.

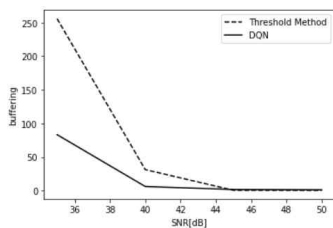


그림 1 재생 끊김 발생 횟수

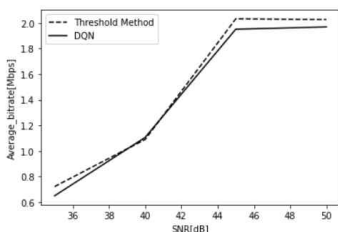


그림 2 비디오 스트림의 평균 비트율

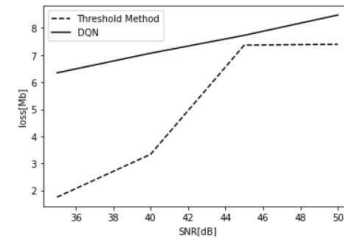


그림 3 전송 손실량

Fig. 1에서 DQN 모델이 모든 SNR에 대해 더 좋은 성능을 보임을 알 수 있다. Fig. 2에서는 두 모델의 평균 비트율이 비슷함을 볼 수 있는데, 여기서 DQN 모델이 평균 비트율은 임계 전송 모델과 근사하지만, 재생 끊김 횟수는 더 좋은 성능을 보이는 것을 확인할 수 있다. 단, Fig. 3에서 DQN 모델의 송신 손실이 더 크게 나타나는 것을 볼 수 있는데, 이를 통해 DQN 모델은 비트율을 낮추어 재생 끊김을 막는 대신에, 전송 손실이 발생하더라도 더 많이 송신하는 방향으로 학습했다고 볼 수 있다.

두 모델의 경향성을 비교했을 때, 재생 끊김 횟수와 전송 손실에서 임계 전송 모델이 SNR의 변화에 대해 더 민감하게 변화한다. 이를 통해 DQN 모델이 임계 전송 모델보다 더 활용성이 있다고 볼 수 있는데, 이는 DQN 모델이 채널 용량 $C(t)$ 를 모르는 상태로 학습했으므로 채널 환경의 변화에 대해 같은 알고리즘으로도 학습 공간을 갱신해 지속적으로 대응할 수 있기 때문이다. 반면, 임계 전송 방식은 SNR이 바뀔 때마다 큐 임계값에 따른 전송 규칙 자체를 매번 바꾸어 주어야 하는 단점을 가진다.

V. 결론

본 논문에서는 채널 정보가 전혀 없는 상황에서 비디오 스트리밍 서비스의 재생 지연을 제한하면서 평균 비트율을 높이는 DQN 기반 SVC 비디오 전송 기술을 제안하였다. 임계값 기반의 결정론적인 방법 대비 더 좋은 성능을 보이면서, 강화학습을 활용하여 사용자 큐잉 시스템을 조절하여 딜레이-민감 서비스를 달성할 수 있음을 보였다.

ACKNOWLEDGMENT

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No. 2021-0-02201, 사용자 프라이버시를 보존하는 비디오 캐싱을 위한 연합 학습 시스템)과 2023년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2022R1C1C1010766, No. 2022R1A4A3033401).

참고 문헌

- [1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2017-2022 White Paper," 2019.
- [2] M. Choi, A. No, M. Ji, and J. Kim, "Markov decision policies for dynamic video delivery in wireless caching networks," IEEE Trans. Wireless Commun., vol. 18, no. 12, pp. 5705 - 5718, Dec. 2019.
- [3] C. Ye, M. C. Gursoy, and S. Velipasalar, "Power control for wireless VBR video streaming: From optimization to reinforcement learning," IEEE Trans. Commun., vol. 67, no. 8, pp. 5629 - 5644, Aug. 2019.
- [4] W. J. Yun, D. Kwon, M. Choi, J. Kim, G. Caire and A. F. Molisch, "Quality-Aware Deep Reinforcement Learning for Streaming in Infrastructure-Assisted Connected Vehicles," IEEE Transactions on Vehicular Technology, vol. 71, no. 2, pp. 2002-2017, Feb. 2022.
- [5] C. Kreuzberger, D. Posch, and H. Hellwagner, "A scalable video coding dataset and toolchain for dynamic adaptive streaming over HTTP," in Proc. ACM Multimedia Syst. Conf., 2015, pp. 213 - 218.