

강화학습 vs. 최적화: 콘텐츠 캐싱 네트워크에의 응용

곽정호

대구경북과학기술원 (DGIST) 정보통신융합전공

요 약

본 논문에서는 지금까지 연구되어 온 여러가지 콘텐츠 캐싱 네트워크 기술들에 대해서 알아본 후, 콘텐츠 캐싱 네트워크 문제를 최적화 관점과 강화학습 관점에서 접근하여 해결책을 제시한 연구들의 장단점들을 비교, 분석해본다.

I. 서 론

최근, UHD 와 같은 초고화질의 비디오 콘텐츠 서비스들이 모바일을 포함한 다양한 하드웨어 플랫폼으로 제공됨에 따라 높은 서비스 품질을 위해 네트워크로 하여금 상당히 낮은 전송지연을 요구하고 있다. 전송지연을 줄이기 위한 방법으로는 여러가지 방법이 있을 수 있지만, 엣지 콘텐츠 캐싱은 그 중 가장 간단하면서도 확실한 방법이라고 할 수 있다.

엣지 콘텐츠 캐싱 기술의 개념은 다음과 같다. 하나의 서버에서 전 세계의 여러 지역에게 비디오 데이터를 전송하기 위해서는 지역적으로 먼 거리에 있으므로 당연히 긴 시간이 필요하다. 그런데 해당 비디오 데이터를 요청하기 이전에 가까운 지역의 서버에 그 데이터를 저장해 놓는다면 물리적인 거리가 짧으므로 전송 지연을 줄일 수 있다.

지금까지 콘텐츠 캐싱 네트워크 연구는 콘텐츠를 전송하기 위해 최초로 한번 원래의 서버로부터 콘텐츠를 받은 후, 해당 콘텐츠를 계속 엣지 서버에 저장할 것인지 혹은 버릴 것인지를 결정하는 연구와 (수동 캐싱) 콘텐츠의 인기도를 빅데이터나 머신러닝 알고리즘 등으로 예측한 후, 인기가 있을 것 같은 콘텐츠들을 미리 해당 엣지 서버에 저장하는 연구 (능동 캐싱)으로 나누어서 진행되어 왔다. 최근에는 능동 캐싱 문제의 솔루션을 최적화 기법과 강화학습 기법으로 찾아내는 연구들이 많아지게 되었다.

II. 본 론

최적화 이론을 활용한 계층적 클라우드 vs. 엣지 캐싱 [1]. 최근 클라우드와 엣지가 공존하는 계층적 구조가 5G 네트워크의 구조로써 제안되고 있다. 이 계층적 구조에서 클라우드에 콘텐츠를

캐싱하는 것의 장점은 큰 스토리지와 넓은 커버리지가 될 수 있고, 엣지에 콘텐츠를 캐싱하는 것의 장점은 낮은 전송지연이 될 수 있다. 따라서 이 연구에서는 이러한 클라우드와 엣지의 계층적 구조를 가지는 시스템에서 평균 전송 지연을 최소화하는 최적화 문제를 만들고, 일반적으로 인기있고, 큰 크기이고, 딜레이에 둔감한 파일을 클라우드 서버에 캐싱하고, 작은 크기이고, 딜레이에 민감한 파일을 엣지 서버에 캐싱하는 최적 해법을 찾아낼 수 있었다.

강화학습을 활용한 계층적 클라우드 vs. 엣지 캐싱 [2]. 시공간으로 변화하는 콘텐츠의 인기도를 예측하는 것은 상당히 어렵다. 이 연구에서는 일반적인 인기도와 지역적인 인기도를 구분하여 인기도의 상태를 고려하여, 클라우드 서버와 엣지 서버에 캐싱될 콘텐츠를 결정하는 것을 Action 으로 하고, 해당 일반적 인기도와 지역적 인기도를 State 로 하는 Markov Decision Process 모델을 만들고, 강화학습 기반의 해법을 제시하였다. 여기에서, 강화학습 문제에서 흔히 발생하는 이슈인 파라미터의 종류가 증가함에 따라 알고리즘 복잡도가 급격하게 증가하는 Curse of Dimensionality 를 해결하기 위한 근사 방법을 적용한 알고리즘을 제안하였다.

III. 결 론

본 논문에서는 최적화 기법과 강화학습 기법을 활용한 콘텐츠 캐싱 네트워크 연구들을 살펴보았다. 강화학습 기법은 최적화 기법에 비해 고려할 수 있는 모델의 자유도가 높은 장점이 있는 반면, 알고리즘의 복잡도가 높은 문제로 인해 아직까지 적은 수의 State 만을 고려한 문제들을 다룸으로써 현실적인 모델링을 잘 반영하지 못한다는 단점이 있다.

참 고 문 헌

- [1] J. Kwak *et al.* "Hybrid Content Caching for 5G Wireless Networks: Cloud versus Edge Caching," IEEE Trans. on Wireless Commun., May 2018.
- [2] A. Sadeghi *et al.* "Optimal and Scalable Caching for 5G using Reinforcement Learning of Space-Time Popularities," IEEE Journal on Selected Topics in Signal Processing, Feb. 2018.