https://doi.org/10.7840/kics.2023.48.3.327

5G 특화망내 심층강화학습 기반 네트워크 캐싱

임 준 영•, 김 동 주*, 유 영 환°

Deep Reinforcement Learning-Based Content Caching for Private 5G Network

Joonyoung Lim•, Dongju Kim*, Younghwan Yoo°

요 약

4차산업 혁명과 함께 5G 특화망의 수요가 증가하였지만 일반적 네트워크를 위한 기존의 운용 기법으론 5G 특화망의 효율적인 관리가 불가능하므로 개별 특화망의 특성에 맞는 운영 시스템이 필요하다. 본 논문에서는 eMBB 서비스를 타겟으로 하는 5G 특화망에서 백홀의 과부하를 줄이고 사용자 QoS를 높이기 위해 심층강화학습기반 캐싱 시스템을 제안한다. 제안하는 시스템은 캐시 배치 단계에서 교체 정책을 고려하는 통합 캐싱 시스템으로서 배치 알고리즘과 교체 정책을 각각 선택하는 기존의 캐싱 전략들과 성능을 비교하였다. 시뮬레이션 결과를 통해 제안하는 캐싱 시스템이 캐시 적중률과 네트워크 지연시간 모두 기존의 전략들보다 20% 이상 우수함을 확인할 수 있었다.

키워드 : 5G 특화망, 심층강화학습, 네트워크 캐싱
Key Words : Local 5G, Private 5G, Deep reinforcement learning, Network cache

ABSTRACT

Although the demands for local 5G network has increase along with the 4th industrial revolution, current network operation techniques and systems cannot efficiently manage local 5G networks, and suitable system for those networks are necessary. Therefore, we propose a deep reinforcement learning based caching system to reduce backhaul overload and increase user QoS in local 5G for efficient network resource utilization in eMBB targeted local 5G networks. The proposed system considers replacement policies in the stage of cache allocation, and its performance is compared with existing caching strategies combined with cache allocation algorithm and cache replacement policy. The simulation result shows the proposed system has 20% higher performance in both cache hit ratio and average network latency than conventional systems.

Ⅰ. 서 론

1.1 시스템 모델 및 문제 구성

최근 사회의 다양한 분야에서 산업의 지능화가 진행되고 있다. 4차 산업화 프로세스 혁명의 스마트 공장, 자율주행 자동차, 스마트시티, 스마트 항구와 같은 시스템의 구축을 위해서 5G 통신의 수요가 증가하였으며 이러한 트랜드는 기업의 5G 인프라 자체 보유로

※ 이 과제는 2022학년도부산대학교교수국외장기파견지원비에 의하여 연구되었음.
• First Author : Pusan National University, School of Computer Science and Engineering, ingyuh1015@pusan.ac.kr, 학생회원
° Corresponding Author : Pusan National University, School of Computer Science and Engineering, ymomo@pusan.ac.kr 종신회원
* Pusan National University, School of Computer Science and Engineering, rlaehdwn9097@pusan.ac.kr
논문번호 : 202212-308-B-RN, Received December 20, 2022; Revised January 21, 2023; Accepted January 31, 2023

이어지게 되고, 5G 특화망의 개념이 탄생했다. 5G 특화망이란 기업이나 대학 등 특정 주체가 5G 네트워크를 직접 구축하거나 이동 통신망 사업자와의 개별 계약을 통해 독립적으로 사용할 수 있는 목적의 사설 네트워크를 뜻한다.

5G 네트워크에서 서비스 유형은 표 1과 같이 uRLLC (Ultra Reliable & Low Latency Communication), eMBB(Enhanced Mobile Broadband), mMTC (Massive Machine Type Communication)로 나뉘는데, 각 서비스들은 성능 요구조건이 각기 다르다. uRLLC는 1ms 수준의 지연시간을 요구하고, eMBB는 최소 10Gbps의 데이터 전송속도, mMTC에서는 면적 당 1백만개의 연결을 지원해야 한다. 각 서비스들의 요구사항이 다르기 때문에 모든 서비스를 동일하게 처리하는 기존의 네트워크 운용으로는 효율적인 서비스를 하기 힘들다.

5G 네트워크에서 서비스의 요구사항을 만족시키기 위해서는 제한된 네트워크 자원의 효율적인 활용이 필요하다. 네트워크 캐싱은 자원을 효율적으로 사용할 수 있는 한가지 방법이다. 해당 네트워크에서 자주 사용되는 콘텐츠를 네트워크 노드, 기지국 등 특정 장소에 저장하여 해당 콘텐츠에 대한 요청이 발생할 경우, 서버단까지 갈 필요 없이 콘텐츠를 저장하고 있는 중간 단에서 요청을 해결할 수 있다. 이를 통해 콘텐츠를 보다 빠르게 전달하여 QoS(Quality of Service) 시간을 만족시킬 수 있을 뿐만 아니라 네트워크의 병목현상을 줄일 수 있다. 또한 네트워크 내 중복 트래픽을 줄임으로써 주파수 대역폭 사용 최적화 또한 가능하다. 기존의 네트워크의 경우 캐시 서버를 두어 네트워크 캐싱을 지원했지만, 네트워크 디바이스들의 발전으로 인해 기지국, 릴레이 노드 또는 엣지 디바이스에도 콘텐츠 저장이 가능해졌고 이를 통해 네트워크 효율을 극대화할 수 있다.

표 1. 5G서비스별 네트워크 요구사항
Table 1. requirements of 5G services

Type of service	Target value
uRLLC	User plane latency: 1ms Control plane latency: 20ms Reliability: 10-5error rate per frame
eMBB	Peak down link data rate: 20 Gbps , Peak up link data rate :10 Gbps User plane latency: 4ms Control plane latency: 10ms
mMTC	Connection: 1,000,000 device/km2

본 논문에서는 5G 특화망의 효율적인 운용을 위하여 심층 강화학습을 활용한 캐싱 전략을 제안하며 이를 eMBB가 주 서비스인 5G 특화망을 시뮬레이터에 구현하여 설명한다. 제안하는 전략은 네트워크의 총 캐시 적중률을 높여 백홀 부하를 줄이고, 네트워크 지연시간을 줄여 QoS를 높이는 것을 목표로 전체 네트워크 내 효율적인 캐시 저장소 관리를 목적으로 한다.

제안하는 캐시 시스템은 심층강화학습 모델을 기반으로 캐시 배치와 교체를 동시에 고려하여 기존의 캐시 알고리즘들에 비해 더 높은 캐시 적중률과 낮은 네트워크 지연을 보였다. 본 연구의 기여는 다음과 같다.

1. 개별 5G 특화망의 핵심 응용의 특성을 고려한 심층강화학습(deep reinforcement learning) 기반 캐시 시스템을 제안한다.
2. 캐시 배치 단계에서부터 향후 캐시 교체 정책을 고려한 네트워크 캐싱 기법을 제안한다.
3. 기존의 교체 및 배치 알고리즘의 조합들의 비교 결과를 통해 특정 네트워크에서 절대적으로 우월한 알고리즘은 없으며 네트워크의 특성을 고려하여 배치 및 교체 알고리즘의 선택이 필수적임을 밝힌다.
4. 제안한 시스템의 심층강화학습의 보상함수를 설계하여 에이전트가 이기종 네트워크 내 최적의 위치에 캐시를 가능하게 한다.

본 논문은 다음과 같이 구성된다. 2장에서는 5G 네트워크에서의 네트워크 캐싱과 관련된 기존 연구와 네트워크 캐싱 문제를 강화학습 알고리즘으로 해결하려는 기존의 연구를 설명한다. 3장에서는 시스템 모델을 소개하고, 문제를 정의한다. 4장에서는 본 논문에서 제안하는 콘텐츠 캐싱 전략에 대해 자세히 설명하고 5장에서는 제안하는 시스템을 구현하여 실험을 진행하며 기존의 캐싱 시스템과 비교 결과를 분석한다. 6장에서는 결론과 향후 연구 방향을 간략히 기술한다.

Ⅱ. 관련 연구

넷플릭스, 구글과 같은 기업들은 자체 네트워크의 효율적 운영과 제공하는 서비스의 지연시간을 줄이며 메인 서버의 부하를 줄이기 위해 콘텐츠 전송 네트워크(contents delivery network)를 사용한다. 콘텐츠 전송 네트워크는 유저에 콘텐츠를 전송하기 위한 네트워크로 실제 서버보다 유저에게 더 가까운 물리적 위치에 구축되어 캐시 서버 역할을 한다. 최근 연구에서는 유저 근처에 캐시 서버를 두는 방식만 아니라, 해

당 유저가 속한 네트워크의 노드를 캐시 저장소로 사용하는 방식에 대한 연구도 진행되었다. 유저가 속한 엣지 네트워크 내 콘텐츠 캐싱은 해당 네트워크의 전체 오버헤드를 감소시키며 콘텐츠 전달 과정의 전체 홉 수를 줄인다. 따라서 전체 네트워크 내 트래픽을 감소시키며 전송 지연을 감소시켜 QoS를 높이는 효과를 보인다. 최근 연구에서는 네트워크 구성 요소인 매크로 셀(macro cell), 스몰 셀(small cell), 그리고 중계 노드에 콘텐츠를 저장하여 캐싱을 지원하는 연구를 진행했다[1] [2][3]. 이외에도 사용자 디바이스[4]와 클라우드 무선 접속 네트워크 (cloud radio access network)[5] 의 원격 무선장비(Remote Radio Head, RRH) 와 중앙집중식 기본대역장치(Base Band Unit, BBU)에 캐싱을 지원하는 연구가 진행되었다[7] [8].

네트워크 캐싱 전략은 크게 캐시 배치(placement) 전략과 캐시 교체(replacement) 정책 두 파트로 나뉘며 선행 연구[8]에서 정리하여 다룬다. 전통적인 캐시 배치 전략으로는 전달 경로 내 모든 노드에 콘텐츠를 저장하는 Leave Copy Everywhere (LCE), 확률적으로 전달 경로 내 랜덤으로 선택된 하나의 노드에 저장하는 Random (RND), 노드가 콘텐츠 요청에 응답할 때마다 요청 사용자에게 가까운 방향으로 한 홉 씩 저장 위치를 이동시키는 Leave Copy Down (LCD) 등이 있다. 전통적인 캐시 교체 정책으로는 사용 횟수가 가장 적은 콘텐츠를 교체하는 Least Frequently Used (LRU), 가장 오랫동안 사용되지 않은 콘텐츠를 교체하는 Least Recently used (LRU) 그리고 가장 먼저 저장된 콘텐츠를 교체하는 First-In First-Out (FIFO) 이 있다.

최적의 캐시 배치 문제는 NP-난해 문제로 알려져 있다. 네트워크 내에 캐시 저장소가 하나라면 인기도와 크기를 고려하여 가장 유리한 콘텐츠를 저장할 수 있지만, 네트워크에 다수의 캐시 저장소가 여러 층에 걸쳐 배치되어 있다면 분산하여 콘텐츠를 저장하는 것이 효율적일 것이다.

NP-난해 문제를 해결하기 위해 자주 사용되는 기법으로 큐러닝(Q-learning) 알고리즘이 있다[9]. 강화학습의 한 종류인 큐러닝은 마르코프 결정 과정(Markov decision process)으로 문제를 정의한 후, 특별한 모델 없이 에이전트가 여러 상황에서 어떤 행동을 해야 최선의 선택일지를 결정하는 최적의 정책을 학습한다. 최근 연구[10]에서 저자는 큐러닝을 사용하여 콘텐츠 캐싱을 하는 클라우드 서비스를 제안했다. 다만 큐러닝은 상태-행동(Q-value)값을 가지는 큐 행렬(Q-table)을 학습에 사용하는데, 상태 공간과 행동 공간이 커지게 되면 큐 행렬의 크기가 기하급수적으로 커져 학습에 필요한 메모리가 증가하고 탐색시간이 증가하며, 큐 행렬에 존재하지 않는 상태일때 어떤 행동을 취해야 할지 결정할 수 없는 문제가 발생한다.

심층강화학습은 이러한 큐러닝의 한계를 해결하기 위해 큐행렬의 값을 심층 신경망(Deep Neural Network, DNN) 모델로 예측하여 모든 상태-행동 값을 저장하지 않는다. 심층강화학습은 각 타임 스텝별로 얻은 경험 샘플들을 버퍼에 저장해 두고, 임의로 뽑아 소형 배치(mini-batch)를 구성해 모델의 학습에 사용한다. 이를 통해 인접한 데이터간 의존도를 낮추고 데이터 활용도를 높일 수 있다. 심층강화학습을 사용하여 모바일 엣지 네트워크에서 캐시 배치 문제를 해결하는 연구는 [11]에서 정리하였다.

Ⅲ. 시스템 모델 및 문제 정의

3.1 시스템 모델

본 논문에서는 다음과 같이 네트워크를 가정한다.

- 각 사용자 디바이스는 하나의 스몰 셀과 연결되어 있다.
- 네트워크의 구성 요소들은 최대 전송 속도로 통신이 가능하다.
- 모든 콘텐츠는 동일한 크기를 가진다.

$$D_{max} = 10^{-6} \sum_{j=1}^{J} \{V_l \cdot Q_l \cdot f \cdot R_m \cdot \frac{12\ N_{BW,\mu}}{T_\mu} \cdot (1 - OH)\} \quad (1)$$

실제 5G 특화망 내에는 여러 종류의 계층이 구현 가능하다. 본 논문에서의 5G 특화망 구조는 그림 1과 같이 public 네트워크/코어 인터넷과 통신 가능한 중앙제어장치, 매크로 셀, 소형 셀 총 3계층 트리 구조와 사용자 디바이스들로 이루어지며 각 노드들은 고유한 id 값을 가진다. 여기서 매크로 셀의 집합은 N $(n_1, n_2, n_3 \ldots n_{max} \in N)$, 소형 셀의 집합은 M $(m_1, m_2, m_3 \ldots m_{max} \in M)$으로 정의한다. 모든 계층의 노드는 캐싱 가능한 공간을 가지며 일정량의 콘텐츠를 저장하고 있다. 특정 사용자가 콘텐츠를 요청할 경우 인접한 소형 셀, 매크로 셀, 중앙제어장치를 포함한 경로를 따라 해당 콘텐츠를 저장하고 있는지 유무를 순차적으로 확인한다. 이때 소형 셀이 요청된 콘텐츠를 가지고 있지 않을 시, 연결되어 있는 매크로

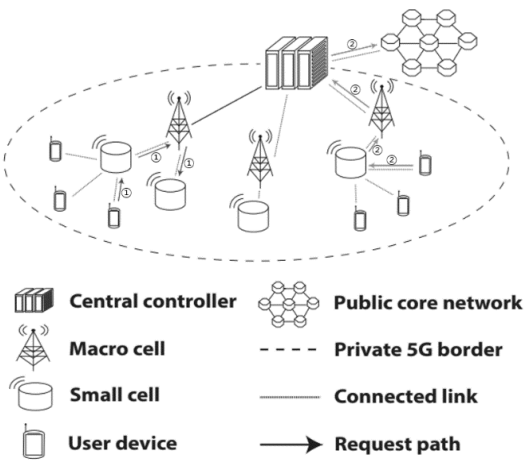

그림 1. 5G 특화망 배치 예시
Fig. 1. Example deployment of private 5G network

셀에 우선적으로 콘텐츠 저장유무를 확인한다. 다만 매크로 셀에 요청 콘텐츠가 존재하지 않을 경우, 바로 연결되어 있는 중앙제어장치에 저장 유무를 확인하는 것이 아니라, 그림 1의 요청 ① 경로와 같이 매크로 셀에 연결되어 있는 다른 소형 셀에 해당 콘텐츠가 저장되어 있는지 확인한다. 연결되어 있는 다른 소형 셀에도 해당 콘텐츠가 저장되어 있지 않을 시 그림 1의 ② 요청 경로와 같이 중앙제어장치의 저장소를 확인하며, 저장되어 있지 않을 경우 공용 네트워크로부터 해당 콘텐츠를 받는다. 만일 경로 내 저장소에 콘텐츠가 저장되어 있을 경우 이를 사용자에게 전송하며 중앙제어장치에 해당 정보를 알리며, 저장하고 있지 않을 경우 공용 네트워크 서버에서 해당 콘텐츠를 다운로드한다.

본 논문에서는 모든 노드들이 충분히 밀접하게 위치하여 최대 전송 속도로 통신이 가능하다고 가정한다. 또한 네트워크 환경이 5G 특화망임을 고려해 통신 속도는 3GPP TS 38.306 표준을 사용한다. 최대 전송 속도 D_{max}는 식(1)을 따르며 파라미터의 값은 표 2를 따른다. 여기서 V_l는 지원하는 MIMO의 레이어 수이며 Q_l는 시스템의 변조 차수이다. f는 스케일 팩터, R_m은 목표 부호화 율을 1024로 나눈 값이다. $N_{BW,\mu}$은 대역폭 BW과 부운반파 간격(sub carrier spacing) μ에 따른 최대 자원 블록 할당 값이다. T_μ는 OFDM의 심볼 지속시간을 뜻하며 OH 는 3GPP에서 지정한 오버헤드 값이다. 이를 통해 통신 대기 시간은 다음과 같이 정의할 수 있다.

$$d_j^i = d_{propa}(i,j) + d_{trans} + d_{queue} \tag{2}$$

$d_{propa}(i,j)$는 전파지연으로 아래와 같다.

$$d_{propa}(i,j) = \frac{dist(i,j)}{l} \tag{3}$$

$dist(i,j)$는 노드 i에서 노드 j 까지의 유클리디안 거리를 뜻하며 l 는 물리적 신호의 속도이다. d_{trans}는 데이터 전송 속도로 다음과 같다.

$$d_{trans} = \frac{s}{D_{max}} \tag{4}$$

s는 패킷의 크기이며 (1)에서 정의한 D_{max}를 사용한다. d_{queue}대기 지연 시간으로 아래와 같이 정의된다.

$$d_{queue} = I \cdot (1-I) \cdot \frac{s}{D_{max}} \tag{5}$$

I는 트래픽 강도(traffic intensity)로 본 논문에서는 0과 1사이의 랜덤 값을 가우시안 분포를 따라 가지도록 한다.

본 논문은 5G의 3가지 시나리오 중 eMBB 시나리오에 집중한다. 높은 대역폭과 더 낮은 지연율을 요구하는 eMBB는 100Mbps의 사용자 체감 속도와 20Gbps의 최대 전송속도 제공을 목표로 하며, 4k 미디어, 증강현실(AR), 가상현실(VR) 등의 어플리케이

표 2. 3GPP TS 38.306 통신 속도
Table 2. 3GPP TS 38.306 data rate

Parameters	description	Values
V_l	Max number of supported layers	4
Q_l	Maximum supported modulation order	6
f	Scaling factor	1
R_m	Value depends on target code rate	0.92578124
BW	Allocated bandwidth	50MHz
μ	Allocated Sub carrier spacing	60khz
$N_{BW,\mu}$	Number of allocated resource block	66
T	Average OFDM symbol duration	$\frac{10^{-3}}{14 \cdot 2^{\mu}}$
OH	Overhead value	0.18

션 서비스가 해당된다. 본 논문에서는 실제 현실과 유사한 환경의 시뮬레이션을 위해 한국 공중파 방송국인 KBS, MBC, SBS 세 방송사의 100여개의 콘텐츠를 생성하여 사용자가 요청할 수 있도록 하였다. 콘텐츠는 뉴스, 예능, 드라마 총 3가지 카테고리가 존재하고, 각 요일 별 시청률을 기반으로 콘텐츠의 요청빈도는 로그 정규분포(log-normal distribution)를 따른다. 이때 대체로 뉴스 카테고리는 일주일 내내 비슷한 요청빈도, 예능과 드라마의 경우 특정 요일에 높은 요청빈도를 띈다.

3.2 문제 정의

eMBB의 경우 8k 품질 혹은 그 이상의 비디오 서비스, AR, VR 서비스 등이 포함된다. 해당 서비스를 타겟으로 하는 특화망의 경우 요청 콘텐츠의 크기가 크기 때문에 코어 네트워크와 연결 링크인 백홀(backhaul)에 과부하가 걸리기 쉽다. 더 많은 백홀의 필요성은 기존 네트워크에서 eMBB 콘텐츠를 서비스할 경우 발생하는 문제 중 하나이며, eMBB 서비스를 타겟으로 하는 5G 특화망을 구축할 경우 고려하여 설계해야 한다.

eMBB서비스를 타겟으로 하는 5G 특화망에서 캐싱 기법을 사용한다면, 백홀을 통해 코어 네트워크에서 전송받아야 할 콘텐츠들을 특화망 내 노드에서 전송을 받을 수 있다. 따라서 백홀의 부하를 줄이는 효과를 기대할 수 있고 이는 사용자 입장에서는 서비스 만족도를 높이는 동시에, 네트워크 운용자의 입장에서는 적은 예산으로 네트워크를 구축할 수 있다.

본 논문에서는 제안하는 캐싱 시스템은 5G 특화망의 캐시 적중률을 높여 백홀 부하를 줄이며 네트워크 지연 시간을 줄여 사용자 서비스 만족도를 높이는 것을 목표로 한다. 캐시 적중률 P_{hit}은 다음과 같이 정의한다.

$$P_{hit} = \frac{\sum_{t=1}^{t_{max}} p(t)}{t_{max}} \tag{6}$$

여기서 $P(t)$는 아래 식을 따른다.

$$p(t) = \begin{cases} 1, & \text{캐시 적중 할 경우} \\ 0, & \text{그렇지 않을 경우} \end{cases} \tag{7}$$

t는 발생한 eMBB 콘텐츠의 요청된 순서를 나타낸 값이며, t_{max}는 네트워크에서 발생한 총 요청 횟수이다.

네트워크 총 지연시간은 다음과 같이 정의한다.

$$d_{total} = \frac{\sum_{t=1}^{t_{max}} d(t)}{t_{max}} \tag{8}$$

여기서 $d(t)$는 다음과 같이 표현된다

$$d(t) = \sum_{i=0}^{len(path(t))-1} d_{path(t)[i]}^{path(t)[i+1]} \tag{9}$$

$path$(t)는 요청 패킷이 전송되는 경로 리스트이다.

따라서 본 논문의 문제 정의는 다음과 같이 표현할 수 있다.

$$\begin{aligned} P1: Maximize \quad & P_{hit} \\ P2: Minimize \quad & d_{total} \end{aligned} \tag{10}$$

P1은 캐시 적중율을 최대화하는 문제이고, **P2**는 전체 네트워크 지연시간을 최소화하는 문제이다.

Ⅳ. 심층 강화학습 기반 캐싱 시스템

기존의 연구에서는 콘텐츠 배치 전략과 교체전략을 따로 적용하여 네트워크 캐싱을 진행하였지만, 본 논문에서는 배치와 교체를 통합한 콘텐츠 캐싱 전략을 목표로 한다.

제안 알고리즘의 목표는 5G 특화망 내 캐시 배치를 통해 캐시적중률(cache hit rate)을 높여 백홀의 부하를 줄이며 최종적으로 네트워크 지연시간을 최소화하여 사용자 만족도를 높이는 것이다. 또한 캐시 배치 결정 단계에서 교체 정책을 고려해 각 정책의 효율을 최대화하는 것을 목표로 한다. 위 목표를 위해 본 논문에서는 심층 신경망을 이용한 강화학습의 일종인

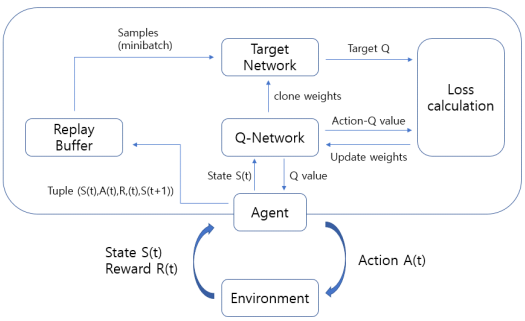

그림 2. 심층 강화학습 개념도
Fig. 2. Architecture of deep reinforcement learning

심층강화학습 모델로 접근한다. 본 논문에서는 구글 딥마인드에서 제안한 심층강화학습 모델을 사용하였다[12]. 그림 2는 실제 구현한 심층강화학습 프레임워크이며 각 요소에 대한 설명은 아래와 같다.

4.1 에이전트와 심층 신경망 모델

에이전트는 중앙제어장치에 위치한다. 사용자 디바이스가 콘텐츠를 요청할 경우 해당 요청은 순차적인 경로를 통해 중앙제어장치에 전달된다. 이때 콘텐츠가 해당 경로에 존재하지 않을 경우 중앙제어장치는 공용네트워크에서 해당 콘텐츠를 다운로드하여 사용자 디바이스에 제공해 주며, 해당 네트워크의 어느 저장소에 콘텐츠를 캐싱 할지 혹은 캐싱을 하지 않을지 행동을 취한다. 에이전트에는 2개의 심층 신경망 모델과 재현 버퍼(replay buffer)가 존재하며, 각각의 심층신경망 모델은 Q-네트워크, 목표 네트워크(Target network)로 불린다.

Q-네트워크는 상태공간을 입력 값으로 받으며 행동 공간의 인자에 해당하는 상태-행동 값을 출력한다. Q-네트워크의 상태-행동 값을 구하는 상태-행동 함수는 식(11)과 같이 표현한다. S(t)는 시간 t에서의 상태 공간을 뜻하며 a_j 는 행동 공간의 임의의 인자이다. θ는 심층 신경망 내 가중치 벡터들의 집합을 뜻하며 λ는 미래 보상의 중요도를 뜻하는 감쇠 인자(discount factor)이다. $R(S(t), a_j)$는 상태 S(t)에서 행동 행동 a_j를 통해 얻는 보상 값이다. 에이전트는 해당 값을 기반으로 가장 높은 보상이 예상되는 행동을 실행한다. 따라서 에이전트의 궁극적인 목표는 상태-행동 값을 최대화하는 최적의 정책을 찾는 것이며 이를 위해 목표 네트워크를 사용한다

목표 네트워크는 주기적으로 Q-네트워크의 가중치를 동일하게 복제하여 이중화 된 구조로 만들며 식(12)와 같은 값을 출력한다. 해트 부호(Hat sign)은 타겟 네트워크의 복제 주기가 돌아오기 전까지 Q-네트워크와 심층신경망 내 가중치의 차이가 존재하기 때문에 이를 구분하기 위해 표시한다. 에이전트는 Q-네트워크의 출력 값 $Q(S(t),a_j;\theta)$와 타겟 네트워크의 출력 값 $y(t)$를 손실함수(loss function)에 대입하며 이를 최소화하는 방향으로 Q-네트워크 모델을 수정해 나가며 평균 제곱 오차(mean squared error)를 따르는

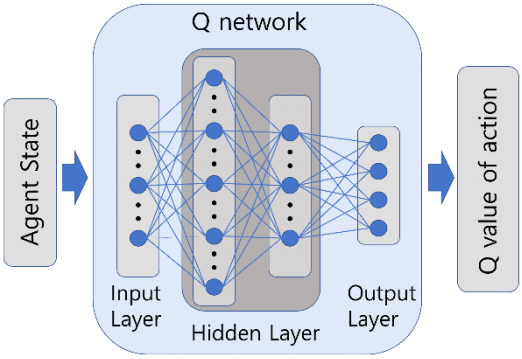

그림 3. 심층 신경망 구조
Fig. 3. Structure of deep neural network model

손실함수는 다음과 같다.

$$L(t) = [y(t) - Q(S(t+1),a_j;\theta)^2] \quad (13)$$

강화학습 에이전트에 적용된 신경망은 심층 신경망으로 그림 3과 같이 입력층과 출력층 사이에 여러 개의 은닉층으로 이루어져 있다. 입력층의 유닛 수는 에이전트의 상태 공간과 동일하게 105개로 구성하였다. 사용한 신경망은 2개의 은닉층을 가지는데, 첫번째 은닉층의 유닛 수는 입력층의 세 배 크기인 315개로 구성되어 있으며 두번째 은닉층의 유닛 수는 입력층과 같은 크기인 105개로 구성되어 있다. 마지막 출력층의 유닛 수는 강화학습 에이전트의 행동 공간의 크기인 4개로 구성되어 있다. 각각의 층들은 완전결합(fully connected) 되어있으며 은닉층에는 과적합을 피하기 위해 드롭 아웃(drop out) 비율을 0.2 로 지정하였다. 입력층과 은닉층의 활성화 함수로는 ReLU(Rectified Linear Unit) 함수를 적용하였으며 마지막 출력층의 활성화 함수로는 Softmax 함수를 적용하였다. 손실 함수로는 평균제곱오차(Mean Squared Error, MSE)를 사용하였으며 최적화 과정에서는 Adam 함수를 적용하였다.

재현 버퍼는 인접한 학습 데이터 사이의 연관성으로 인해 발생하는 문제를 완화하기 위한 기법이다. 각각의 행동에 따른 상태공간 변화를 튜플 형태로 저장하며 각 단계마다 랜덤으로 소형 배치(mini batch) 크기만큼 Q-네트워크와 목표 네트워크에 학습을 위해

$$Q(S(t),a_j;\theta) = R(S(t),a_j) + \lambda \cdot \max Q(S(t+1),a_j;\theta') \quad (11)$$

$$y(t) = \hat{r}(t) + \lambda \cdot \max \hat{Q}(S(t+1),a_j;\hat{\theta}) \quad (12)$$

제공한다. 학습에는 ε-탐욕(ε-greedy) 알고리즘이 사용된다. ε-탐욕 알고리즘은 학습초기에는 ε 값이 1로 설정되어 에이전트가 100% 무작위로 행동하지만, 학습이 진행됨에 따라 점진적으로 ε 값을 지정한 값까지 줄여 나가며 에이전트가 판단하는 비중을 늘려 나간다. 학습에 사용된 각 인자의 값들은 표 3을 따른다.

4.2 상태 공간

상태 공간 S(t)는 다음과 같이 정의된다.

$$\mathrm{S(t)} = \left[C_{req}(t), T(t)\right] \tag{14}$$

$C_{req}(t)$는 t 번째 요청된 콘텐츠에 대한 정보로 다음과 같이 구성된다.

$$C_{req}(t) = [\nu(t), \mu(\nu(t), t)] \tag{15}$$

여기서 $\nu(t)$는 요청된 콘텐츠의 id값 그리고 $\mu(\nu(t), t)$는 콘텐츠의 캐시 교체 정책에 따른 캐시 이득(cache gain) 값이다.T(t)는 요청의 코어 네트워크까지 경로의 저장소에 대한 정보로 다음과 같이 정의된다.

$$T(t) = [X_n(t), X_m(t), X_c(t)] \tag{16}$$

$$X(t) = [C_1(t), \ldots C_{max}(t)] \tag{17}$$

X(t)는 노드의 저장소에 관한 정보로서 해당 저장소에 저장되어 있는 모든 콘텐츠에 대해 식(15)에서 정의된 $C_{req}(t)$와 같은 구조의 값을 담고 있다. 아래 첨자 n, m, c는 각각 $path(t)$의 인자인 스몰 셀, 매크로 셀, 중앙제어장치의 고유 id값을 뜻한다.

4.3 행동 공간

행동 공간의 크기를 제한하기 위해, 본 논문의 에이전트는 콘텐츠 요청이 통과한 경로에만 해당 콘텐츠를 캐싱 할 수 있도록 한다. 행동 공간 A(t)은 다음과 같이 정의된다.

$$\mathrm{A(t)} \in \{a_0, a_1, a_2, a_3\} \tag{18}$$

표 3. 심층강화학습 학습 파라미터 값
Table 3. value of DQN parameters

Type of parameter	Value
Final value of ε	0.2
Decay value of ε	0.99995
Replay buffer size	100000
Learning rate	0.0001
Mini batch size	1024
Decay value	0.95

에이전트는 캐싱을 하거나 경우에 따라 어느 곳에도 캐싱을 하지 않는 선택을 할 수 있다. 따라서 a_0, a_1, a_2는 각각 소형 셀, 매크로 셀, 중앙제어장치에 저장하는 행동을 뜻하고, a_3은 어디에도 저장하지 않는 행동을 뜻한다.

4.4 보상

심층강화학습 모델의 보상 함수는 시스템의 목표를 반영해야 한다. 본 논문에서의 목표는 이전 섹션에서 정의한 **P1, P2**를 통한 캐시 배치 문제와 캐시 교체 문제의 통합적 접근이므로 이를 반영해 수식 (19)와 같이 정의한다. 수식 (19-a)의 경우 에이전트가 저장하는 행동 a_0, a_1, a_2중 하나를 선택하여 해당 콘텐츠가 선택한 네트워크 노드 저장소에 저장이 되어 상태 공간에 변화가 생긴 경우이다. $G(t)$는 선택한 노드에 저장을 함으로써 콘텐츠를 제공할 수 있는 사용자 디바이스의 수이다. 해당항은 본 논문에서 지정한 **P1**, 즉 캐시 적중율을 최대화하기 위한 항이다. 다만 **P1**만 고려한다면 에이전트는 가장 많은 사용자 디바이스에게 콘텐츠를 제공할 수 있는 중앙제어장치에 집중적으로 캐싱을 할 것이다. d_{gain}은 콘텐츠를 저장함으로써 이후에 동일한 요청이 발생할 경우 얻게 되는 네트워크 지연 이득이며 다음과 같이 정의된다.

$$d_{gain} = d_{core} - d_{cache} \tag{20}$$

d_{core}는 코어 네트워크에서 요청된 콘텐츠를 받아올 경우 소요되는 네트워크 지연 시간이며 d_{cache}는 해당 액션을 통해서 저장한 캐시에서 콘텐츠를 전송받는 경우 소요되는 네트워크 지연 시간이다. d_{gain}은 **P2**, 즉

$$R(t) = \begin{cases} \alpha \cdot \log(G(t)) + \beta \cdot d_{gain}(t) & \textit{상태공간이 변경될 경우} \quad \text{(19-a)} \\ R_p & \textit{행동 } a_3 \textit{ 일 때, 상태공간이 변경될 수 없었을 경우} \quad \text{(19-b)} \\ R_n & \textit{그 외 나머지 경우} \quad \text{(19-c)} \end{cases}$$

전체 지연시간을 최소화하기 위한 항이며 사용자 디바이스와 더 가까운 스몰 셀에 저장을 할 경우 더 높은 보상을 얻을 수 있다. 따라서 $G(t)$과 d_{gain}는 서로 트레이드 오프(trade-off) 관계이며 적절한 비율을 찾는 것이 중요하다. 다음 장에서 보상 함수식 (19-a)의 상수 계수 α, β의 최적의 값을 실험을 통해 구한다. 보상 함수 식 (19-b)는 현재 요청된 콘텐츠가 전달되는 경로의 어떤 저장소에도 해당 콘텐츠보다 캐시 이득이 낮은 콘텐츠가 저장되어 있지 않은 경우이다. 어느 저장소를 선택해도 어차피 캐시 교체가 일어나지 않기 때문에 애초에 아무 곳에도 저장하지 않는 행동 a_3을 선택함으로써 양의 보상을 얻을 수 있다. 수식 (19-c)는 나머지의 경우에 음의 보상을 주기 위함이다. 해당 경로에 저장할 공간이 있음에도 저장하지 않을 경우, 특정 저장소에 교체할 콘텐츠가 있음에도 해당 노드를 선택하지 않고 저장할 수 없는 노드를 선택하거나 행동 a_3을 선택하여 저장하지 않는 경우 등이 이에 해당한다. (19-b), (19-c)를 통해 본 논문에서 제안하는 캐싱 시스템이 캐시 배치 문제와 캐시 교체 전략을 동시에 고려하도록 한다.

4.5 교체 정책

본 논문에서 제안하는 시스템에서는 전통적인 캐시 교체 정책 중 하나인 Least Frequently Used(LFU)를 변형하여 특정 기간 동안 네트워크에 요청된 빈도를 기준으로 교체 여부를 결정한다. 기존 LFU의 경우 캐시 배치 알고리즘에서 선택한 위치에 확정적으로 저장을 할 수 있지만 해당 교체 정책의 경우 요청된 콘텐츠의 최근 요청 빈도를 캐시 이득으로 정하여 기존 저장되어 있는 콘텐츠과 캐시 이득을 비교하여 저장 및 교체 여부를 결정한다. 특정 시간 t에 $\lambda(t)$에 해당하는 콘텐츠의 캐시 이득은 $\mu(\lambda(t), t)$와 같이 표기한다.

V. 실험 및 결과

본 논문의 실험은 두 단계로 수행된다. 첫번째는 최적의 보상함수 계수 α, β 를 찾기 위한 실험이고, 두번째는 앞서 찾은 최적의 보상함수로 학습된 모델과 기존의 캐시 배치 및 교체 알고리즘들과 비교하는 실험이다. 비교 실험의 캐시 배치 알고리즘은 LCE, LCD 그리고 RND (Random)알고리즘이 사용되었고 교체 전략으로는 LFU, LRU 와 FIFO가 사용되었다.

논문의 실험은 파이썬 버전 3.10 기반의 시뮬레이터에서 진행되었으며, 학습과 실험이 진행된 장비의 사양은 표 4와 같다. 공통적인 시뮬레이션 파라미터는

표 4. 학습 및 실험에 사용된 장비의 사양
Table 4. value of simulation parameter

CPU	i9-12900KS
GPU	Geforce RTX 3090 Ti
RAM	128GB

표 5. 시뮬레이션 파라미터 값
Table 5. value of simulation parameter

Type of parameter	Value
Field size	1km2
Number of central controller	1
Number of macro cell	4
Number of small cell	9
Packet size	1500 bytes
Simulation round/request	50000
Simulation week	10

표 5를 따른다. 총 50000개의 사용자 요청이 시뮬레이션 내 10주의 기간동안 요청되었고 하나의 사용자 요청은 시뮬레이션의 한 라운드로 정의하였다. 사용자의 요청은 3.1절 시스템 모델에서 기술한 한국 공중파 3사의 시청률을 기반으로 한 시나리오를 사용하여 생성한다. 모든 실험은 10회를 실행하여 평균을 낸 결과이다. 시뮬레이션 내 네트워크는 $1km^2$의 공간에 9개의 스몰 셀, 4개의 매크로 셀, 1개의 중앙제어장치가 존재한다. 각각의 콘텐츠는 소형 셀과 매크로 셀의 저장소 용량의 20%인 고정된 크기를 가지며, 중앙제어장치는 다른 노드에 비해 2배 용량의 저장소를 가진다.

5.1 심층강화학습 보상 함수 실험

본 실험은 식 (19-a)의 최적의 계수 α, β를 찾는 실험이므로 상수 계수의 값을 바꿔가며 실험을 진행하면서 학습된 행동 선택 비중, 캐시 적중률, 평균 통신 홉 수 그리고 평균 네트워크 지연시간을 비교한다. 캐시 적중률과 총 네트워크 지연시간은 3.2장에서 정의한 식 (6)과 (8)을 따르며, 평균 통신 홉은 사용자 디바이스에서 요청한 콘텐츠의 위치까지의 홉 수로 계산하였다. 이때 중앙 제어장치에서 공용 코어 네트워크까지의 홉 수는 1로 가정하였다. 실험에 사용된 α, β의 케이스는 표 6과 같다.

그림 4는 케이스 별로 에이전트가 선택한 전체 행동 수 분포를 나타낸다. 모든 케이스에 있어서 a_3가 가장 많은 분포를 보였으며 이는 시뮬레이션이 진행될수록, 캐시 적중이 되지 않은 콘텐츠 요청의 경우

표 6. 각 케이스 별 상수 계수값
Table 6. value of constant coefficient for each case

Type of parameter	α	β
Case 1	1	1000
Case 2	1	2000
Case 3	2	1000
Case 4	3	1000
Case 5	4	1000

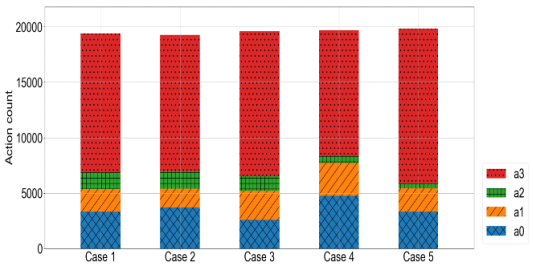

그림 4. 케이스별 각 행동 선택 수
Fig. 4. Number of actions chosen in each case

이미 저장되어 있는 콘텐츠들에 비해 캐시 이득이 낮기 때문에 에이전트가 캐시 교체를 하지 않는 상황이 발생하기 때문이다. 최소 60% 정도를 차지하는 행동 a_3를 제외하고 나머지 행동들을 비교해보면, 케이스 2에서는 a_0이 가장 많은 비중을, 그리고 a_1, a_2가 유사한 비중을 차지한다. 케이스 4의 경우 a_1이 가장 높은 비중을 보였으며 a_0, a_2가 그 뒤를 이었다. 다만 케이스 4, 5의 경우 중앙 제어장치에 저장을 하는 a_2의 선택비율이 전체 행동 수의 4% 이하이며, 이는 저장소가 비어 있을 때 이외에 다른 상황에서 콘텐츠 교체를 고려하지 않았음을 의미한다. 따라서 저장소의 활용도 측면에서 케이스 1, 2, 3이 4, 5에 비해 높음을 알 수 있다.

그림 5는 각 케이스별 평균 캐시 적중률, 평균 홉 수, 평균 네트워크 지연시간 결과이다. 평균 캐시 적중률 결과에 해당하는 그림 5 a)에서 케이스 1이 가장 높은 캐시 적중률을 보이며 그 뒤를 2, 4, 3, 5가 잇는다. 평균 홉 수에 해당하는 그림 5 b)의 경우에 케이스 3이 가장 짧은 홉 통신을 하였음을 확인할 수 있고 1, 2, 5, 4의 순으로 결과를 보인다. 본 논문에서 사용한 네트워크 모델에서는 사용자 디바이스에서 요청이 발생할 경우 부모 노드로만 요청이 진행되지 않고, 매크로 셀 하위에 연결되어 있는 모든 소형 셀을 탐색하며 상위 노드로 요청이 전달되기 때문에 비교적 평균 홉 수가 높은 값을 보인다. 예를 들어 전통적인 트리 네

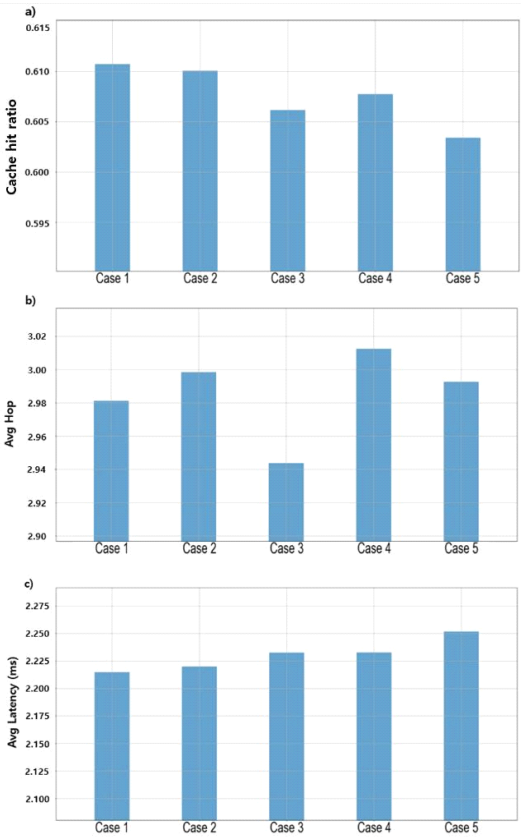

그림 5. 케이스별 성능 비교
Fig. 5. Performance comparison between different reward function coefficient cases

트워크의 경우 소형 셀에서 캐시 적중이 일어날 경우 이는 홉 수가 1이지만, 본 논문에서 사용하는 네트워크에서는 1 또는 3을 가지기 때문에 전체적인 평균 홉 수 또한 높아진다. 네트워크 지연시간 결과인 그림 5 c)의 경우 케이스 1이 가장 짧은 지연 시간을 보이며 캐시 적중률과 비슷한 순서를 보이지만 케이스 3과 4가 거의 동일한 점이 특징이다. 이는 케이스 3이 캐시 적중률은 케이스 4보다 낮지만 평균 통신 홉 수가 케이스 4보다 짧아 서로 상쇄되어 나온 결과이다.

본 논문에서는 3.2에서 정의한 P1과 P2를 고려하였을 때 평균 캐시 적중률 실험과 네트워크 지연 실험에서 가장 좋은 성능을 보였던 케이스 2의 상수 계수가 최적이라 판단하여 캐시 알고리즘 성능 비교 실험에는 케이스 2의 결과를 사용한다. 다만 케이스 2 이외의 다른 케이스들의 캐시적중률이 모두 60% 이상이며 편차가 1% 내외이며, 평균 네트워크 지연시간 편차 또한 0.05 ms 이내의 결과를 보이므로 최적의 계수가 아니어도 네트워크의 성능 향상이 일정 수준

보장되는 것을 확인하였다.

5.2 캐싱 알고리즘 성능 비교

이 절에서는 5.1절에서 도출한 최적의 심층강화학습 기반 캐싱 전략과 다양한 기존의 캐싱 전략의 성능을 비교한다. 제안하는 전략은 이후 그림에서 DQN으로 표기되며, 캐시 배치 전략 3개와 캐시 교체 정책 3개의 총 9가지 조합이 비교 실험에 사용된다. 5.1절과 동일한 3가지 실험의 결과가 소개된다.

그림 6은 제안 기법과 기존 전략들의 3가지 비교실험 비교 결과이며 그림 6의 a), b), c)는 각각 캐시 적중률, 평균 홉 수 그리고 네트워크 지연 시간 결과이다. 제안 기법이 평균 홉 결과는 (LCD, LRU), (LCE, LFU), (LCD, FIFO), (RND, LFU), (LCD, LFU) 조합에 이어 6번째로 짧았지만, 캐시 적중률은 약 62%, 평균 지연시간은 약 2.2ms로 타 조합에 비해 적게는 20% 많게는 50% 더 좋은 성능을 보인다.

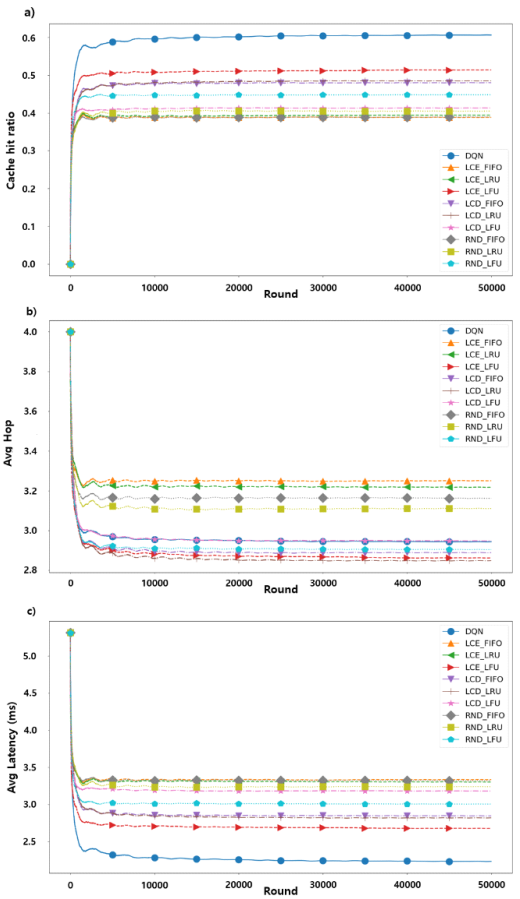

그림 6. 알고리즘별 성능 비교
Fig. 6. Performance comparison with conventional algorithms

비교군의 결과를 통해 캐시 배치 전략이나 캐시 교체 정책 단일 알고리즘으로는 타 알고리즘보다 우월하지 않고 협력 작용이 중요함을 확인할 수 있다. 따라서 네트워크의 특성을 고려하여 맞춤형 배치 전략과 교체 정책 선택이 필수적이다.

Ⅵ. 결 론

본 논문에서는 5G 특화망 환경에서 백홀의 부하를 줄이며, 사용자 QoS 향상을 위해 캐시 적중률과 평균 네트워크 지연 시간을 함께 고려하였다. 이를 효율적으로 처리하기 위해 심층강화학습 기반 모델을 사용하며 캐시 배치 단계에서 교체 정책을 고려하는 시스템을 제안하였다. 실험을 통해 먼저 제안 시스템의 모델 학습의 방향을 결정하는 보상함수 계수의 최적값을 찾은 후, 기존의 다양한 캐시 배치 및 교체 알고리즘들 의 조합과의 비교를 통해 제안한 시스템이 20%~50% 더 우수한 성능을 보임을 확인했다.

본 논문의 실험결과에서 캐시 적중률과 네트워크 평균 지연시간의 순위는 대체로 동조화(coupling)하는 모습을 보이지만 평균 홉 결과는 그렇지 않는다. 이는 실제 네트워크의 경우 공용 코어 네트워크에서도 여러 홉을 통해 요청한 콘텐츠를 전송받지만 시뮬레이션 환경을 구성할 때 공용 코어 네트워크는 별다른 구현 없이 일괄적이게 1홉으로 가정한 것이 원인이라 생각된다. 따라서 향후 연구에서는 5G 특화망뿐만 아니라 공용 네트워크도 신중하게 설계할 필요성이 있다. 본 연구의 결과를 바탕으로 uRLLC와 mMTC 서비스 목적의 5G 특화망을 위한 캐싱 시스템을 연구할 계획이다.

References

[1] D. Liu and C. Yang, "Energy efficiency of downlink networks with caching at base stations," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 907-922, Apr. 2016. (https://doi.org/10.1109/jsac.2016.2549398)

[2] C. Yang, Y. Yao, Z. Chen, and B. Xia, "Analysis on cache-enabled wireless heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. pp. 131-145, Jan. 2016. (https://doi.org/10.1109/twc.2015.2468220)

[3] S.-J. Cao, et al., “Coded caching for relay networks: The impact of caching memories,” *2020 IEEE ITW*, 2021. (https://doi.org/10.1109/itw46852.2021.9457581)

[4] B. Chen, C. Yang, and A. F. Molisch, “Cache-enabled device-to-device communications: Offloading gain and energy cost,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4519-4536, 2017. (https://doi.org/10.1109/twc.2017.2699631)

[5] T. X. Tran, A. Hajisami, and D. Pompili, “Cooperative hierarchical caching in 5G cloud radio access networks,” *IEEE Network*, vol. 31, no. 4, pp. 35-41, 2017. (https://doi.org/10.1109/mnet.2017.1600307)

[6] J. Yao and N. Ansari, “Joint content placement and storage allocation in C-RANs for IoT sensing service,” *IEEE Internet of Things J.*, vol. 6, no. 1, pp. 1060-1067, 2018. (https://doi.org/10.1109/jiot.2018.2866947)

[7] A. Ndikumana and C. S. Hong, “Cooperative and weighted proportional cache allocation for EPC and C-RAN,” in *Proc. Korean Inf. Sci. Soc. Conf.*, pp. 1160-1162, 2017.

[8] W. Ali, S. M. Shamsuddin, and A. S. Ismail, “A survey of Web caching and prefetching,” *Int. J. Adv. Soft Comput. Appl.*, vol. 3, no. 1, pp. 18-44, 2011.

[9] C. J. C. H. Watkins and P. Dayan, “Q-learning,” *Mach. Learn.*, vol. 8, no. 3, pp. 279-292, 1992.

[10] H. Chen and G. Tan, “A Q-learning-based network content caching method,” *EURASIP J. Wireless Commun. and Netw.*, vol. 2018, no. 1, pp. 1-10, 2018. (https://doi.org/10.1186/s13638-018-1268-1)

[11] H. Zhu, et al., “Deep reinforcement learning for mobile edge caching: Review, new features, and open issues,” *IEEE Network*, vol. 32, no. 6, pp. 50-57, 2018. (https://doi.org/10.1109/mnet.2018.1800109)

[12] V. Mnih, et al., “Playing atari with deep reinforcement learning,” *arXiv preprint arXiv:1312.5602*, 2013. (https://doi.org/10.48550/arXiv.1312.5602)

임 준 영 (Joonyoung Lim)

2020년 2월 : 부산대학교 정보컴퓨터공학과 학사

2020년 3월~현재 : 부산대학교 정보융합공학과 석박사통합과정

<관심분야> 이동통신, 무선네트워크, 5G 특화망

[ORCID:0000-0002-4384-839X]

김 동 주 (Dongju Kim)

2022년 2월 : 부산대학교 정보컴퓨터공학과 학사

2022년 3월~현재 : 부산대학교 정보융합공학과 석사과정

<관심분야> 무선 네트워크, 인공지능, 강화학습

[ORCID:0000-0002-2735-9449]

유 영 환 (Younghwan Yoo)

2004년 2월 : 서울대학교 전기컴퓨터공학부 박사

2004년 5월~2006년 12월 : 신시내티대학교, 모바일및분산컴퓨팅센터(CMDC) 연구원

2007년 3월~현재 : 부산대학교 정보컴퓨터공학부 교수

<관심분야> 이동통신, 무선네트워크, 사이버물리시스템, 양자인터넷

[ORCID:0000-0002-2813-6116]