

비용 효율적인 OTT 서비스를 위한 심층강화학습 기반 클라우드 자원 할당 연구

이 예 슬¹ · 장 혜 림¹ · 오 수 빈¹ · 이 연 수¹ · 윤 용 익² · 엄 태 원^{3*}

¹숙명여자대학교 빅데이터분석융합학과 석사과정

²숙명여자대학교 빅데이터분석융합학과 정교수

^{3*}전남대학교 데이터사이언스대학원 부교수

Deep Reinforcement Learning-Based Cloud Resource Allocation for Cost-Effective Over-the-Top Services

Yeseul Lee¹ · Hyerim Jang¹ · Soobeen Oh¹ · Yeonsu Lee¹ · Yongik Yoon² · Tai-Won Um^{3*}

¹Master Course, Department of Bigdata Analysis & Convergence, Sookmyung Women's University, Cheongparo 47 gil, 100, Seoul, Korea(Rep.of)

²Professor, Department of Bigdata Analysis & Convergence, Sookmyung Women's University, Cheongparo 47 gil, 100, Seoul, Korea(Rep.of)

^{3*}Associate Professor, Graduate School of Data Science, Chonnam National University, 61186 Gwangju, Korea(Rep.of)

[요 약]

OTT(Over-The-Top) 서비스 요청이 급증하는 상황에서도 OTT 서비스 플랫폼은 시청자에게 안정적으로 고품질의 비디오 스트리밍 서비스를 제공할 수 있어야 한다. 클라우드 컴퓨팅 기반의 OTT 서비스 플랫폼은 서비스에 필요한 자원을 동적으로 할당할 수 있다는 장점이 있지만, OTT 서비스 수요 변동이 심한 상황에서 클라우드 자원을 적시에 할당하면서 OTT 서비스 품질을 만족시키는 데에는 한계가 있다. 본 논문은 과거의 OTT 시청 데이터로 훈련된 LSTM 기반 심층강화학습 모델을 이용하여 할당 및 해제하는 클라우드 자원의 양을 적응적으로 조정하는 클라우드 기반의 새로운 OTT 자원 관리 방법을 제안한다. 심층강화학습 기반 OTT 자원 할당 시뮬레이션을 통해 제안 방법이 클라우드 자원 낭비를 최소화하면서도 서비스 품질을 유지할 수 있음을 검증하였다.

[Abstract]

Even in a situation where Over-The-Top (OTT) service requests are rapidly increased, an OTT service platform must be able to reliably provide high-quality video streaming services to users. Although a cloud computing-based OTT service platform has the advantage of being able to dynamically allocate the resources required for the services, there is a limitation in satisfying the OTT service quality while allocating cloud resources in a timely manner in a situation where the demand for OTT services is fluctuated significantly. This paper proposes a new cloud-based OTT resource management method that adaptively adjusts the amount of allocated and released cloud resources using an LSTM-based deep reinforcement learning model trained with past OTT viewing data. Through OTT resource allocation simulation based on the deep reinforcement learning, it was verified that the proposed method can maintain the service quality while minimizing cloud resource waste.

색인어 : OTT, 미디어서비스, 클라우드, 자원할당, 강화학습

Keyword : OTT, Media Service, Cloud, Resource Allocation, Reinforcement Learning

<http://dx.doi.org/10.9728/dcs.2022.23.3.531>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 22 January 2022; **Revised** 17 March 2022

Accepted 17 March 2022

***Corresponding Author, Tai-Won Um**

Tel: 

E-mail: stwum@jnu.ac.kr

I. 서론

클라우드 기술의 발전과 인터넷 사용의 확대로 미디어 서비스 영역이 확장되고 있고, 미디어 서비스의 유통과 전송이 디지털화되면서 그 규모가 커지고 있다. 이렇게 멀티미디어 서비스가 활성화되면서 최근에는 OTT (Over-The-Top) 플랫폼 사용자가 급증하고 있다. OTT 플랫폼은 인터넷을 통해 방송 프로그램, 영화 등 미디어 콘텐츠를 제공하는 서비스이다. OTT가 성장할 수 있었던 배경에는 스마트폰의 발달과 보급률 증가, 인터넷 통신망의 발달이 있으며, 특히 COVID-19로 인해 이동제한과 재택근무가 시행되면서 전세계 SVoD (Subscription Video on Demand) 구독자 수가 빠르게 증가하고 있다[1].

OTT 시장규모가 확대됨에 따라 만족도 높은 서비스를 제공하기 위해서는 기술적인 조건이 충족되어야 한다. 사용자가 증가하면 트래픽이 폭증하고 컴퓨팅 부하가 증가하기 때문에 대규모 서비스에 적합한 컴퓨팅 자원을 갖추고 원활한 서비스를 사용할 수 있도록 클라우드 인프라가 지원되어야 한다[2]. 전세계적으로 이용되는 OTT 플랫폼은 서버와 먼 거리에 있는 사용자들에게도 안정적으로 고품질의 미디어 스트림을 전달할 수 있는 콘텐츠 전달 네트워크 및 컴퓨팅 인프라가 요구된다[3]. 그러나 OTT 서비스 제공자가 자체적으로 광범위한 콘텐츠 전달 네트워크와 콘텐츠 스토리지 및 스트리밍 서버 등을 구축하는데 많은 비용과 시간이 들기 때문에, 최근 대부분의 OTT 서비스 제공자들은 클라우드 사업자가 대량의 서버와 스토리지 및 네트워크 장비를 구축하여 대여해주는 클라우드 플랫폼을 사용하고 있다. 클라우드를 이용함으로써 OTT 서비스 제공자는 클라우드를 사용한 만큼만 비용으로 지불할 수 있어서 서비스 인프라 구축 비용을 절감할 수 있고 효율성과 생산성을 증대시킬 수 있으며[4], 빠르게 변화하는 OTT 비즈니스 환경에도 민첩하고 유연한 대응이 가능하다는 장점이 있다.

한편, OTT 플랫폼은 사용자들의 시청패턴 및 콘텐츠의 특성 등으로 인하여 유동적인 사용량의 특성을 보인다. 만약 OTT 서비스 제공자가 공급의 측면에서 충분한 클라우드 자원을 확보하지 못한다면, 서비스 이용 폭주 시간대에 콘텐츠 재생이 끊기거나 체감 품질 (Quality of Experience: QoE) 을 충족시키지 못하는 현상들이 발생할 수 있으며, 이는 OTT 서비스 신뢰도 저하로 직결될 수 있다. 만일 준비한 자원이 많이 남는다면 유휴상태로 인한 비용 손실로 이어질 것이고, 반대로 준비한 자원이 모자라다면 이는 OTT 사업자와 사용자 사이에 존재하는 Service Level Agreement (SLA) 계약에 위반되기 때문에 큰 손실 비용이 발생할 것이다[5].

따라서 OTT 서비스 제공자는 유동적인 서비스 요청 상황을 미리 예측하고 빠르게 대처하여 유휴자원에 대한 비용 손실과 SLA 위반을 예방해야 할 것이다. 즉 사용자의 요구량이 증가하면 OTT 서비스 제공자는 클라우드 자원을 추가적으로 확보하여 끊김 없는 고품질 비디오 시청이 가능하도록 사전에 준비해야 하며, 반면에 시청률이 감소하면 할당받은 클라우드

자원을 해제하는 방식으로 자원을 낭비하지 않고 클라우드 임대 비용을 최소화하여 OTT 서비스 이득을 최대화해야 한다.

이 연구는 Long Short-Term Memory (LSTM) 기반의 심층 강화학습을 기반으로 OTT 서비스 요청이 변화되는 패턴을 학습한 후, 서비스 요청의 변화에 대응하여 선제적으로 클라우드 자원을 할당 및 반환함으로써, OTT 서비스 체감 품질을 유지하면서도 클라우드 임대 비용 감소를 통해 수익을 높일 수 있는 OTT 자원 관리 방식을 제안한다.

1장 서론에 이어서, 본 논문의 2장에서는 미디어 클라우드의 자원 할당 정책 및 강화학습 기반 자원 관리 기술과 관련된 선행연구와 이론적 배경을 살펴보고, 3장에서는 클라우드 기반의 OTT 플랫폼 구조와 심층강화학습 기반의 OTT 자원 관리 알고리즘에 대해 제안한다. 또한, 4장에서는 제안 알고리즘에 대한 시뮬레이션과 성능 분석 결과에 대해 기술하며, 5장에서 본 논문의 결론을 맺는다.

II. 관련 연구

2-1 OTT 서비스를 위한 클라우드 자원 할당 정책

클라우드 자원 할당 정책은 사전예측(proactive) 프로비저닝과 반응적(reactive) 프로비저닝 방법으로 구분된다[6]. 사전예측 프로비저닝은 OTT 사용자들의 자원 요구량을 미리 예측하여 사전에 장기적으로 필요한 자원을 준비하는 방법이며, 반응적 프로비저닝은 서버 모니터링을 통해 자원 요구량이 증가되면 즉시 자원을 할당하는 방법이다.

대표적인 클라우드 서비스인 AWS (Amazon Web Services)에서는 장기적으로 사용할 자원을 사전에 할당받는 예약 인스턴스 방식과 단기적으로 필요한 자원을 요구에 따라 즉각 할당받는 온디맨드 인스턴스 방식을 제공한다[7]. AWS를 비롯한 대부분의 클라우드 서비스에서 예약 인스턴스 방식이 온디맨드 인스턴스 방식이 비해 시간 단위당 인스턴스 임대 비용이 상대적으로 저렴하다[8]. 이에 사용자의 요구량 패턴을 정확히 예측하여 온디맨드 인스턴스보다 예약 인스턴스를 최대한 사용하여 OTT 서비스 운영 비용을 줄이고자 하는 연구가 활발히 진행되고 있다[9][10].

클라우드 자원 비용을 줄이기 위한 사례로, VoD (Video on Demand)를 위한 최적화된 클라우드 자동 확장에 관한 연구에서는 사용자 요청을 실행할 온라인 VM (Virtual Machine) 및 스토리지 유닛을 최적화하는 AMATO (Auto-scaling Movie Allocation and Traffic Optimization) 알고리즘을 적용하여 VM 개수를 최소화하는 방법을 제안하였다[11]. 또한, 클라우드 서비스 제공자의 대역폭 활용률과 사용자에게 영향을 미칠 서비스 품질(Quality of Service: QoS)을 동시에 고려한 효율적인 클라우드 가격 책정 및 결제 전략에 대한 연구가 수행되었다[12].

한편, OTT 사업자가 준비한 자원의 양과 사용자가 요구하

는 자원 양의 차이에 따라 오버 프로비저닝과 언더 프로비저닝 현상이 발생한다. OTT 사업자가 준비한 자원의 양이 사용자 요구량보다 많은 경우 오버 프로비저닝이 발생하게 되는데, 이때 OTT 사업자는 유희자원에 따른 기회비용이 발생하게 된다. 반면 OTT 사업자가 준비한 자원의 양이 사용자가 요구한 자원의 양보다 적은 경우, 혹은 사용량이 급증하는 경우 언더 프로비저닝이 발생하여 자원을 추가로 준비하기 위한 지연시간이 발생한다. 이러한 지연시간 발생은 사용자들의 QoS를 감소시켜 계약 취소 및 SLA 위반에 따른 패널티 비용을 지출하게 된다[13]. SLA 위반을 방지하고, OTT 서비스 사용자들의 품질 향상을 보장하기 위하여, 선행연구에서는 예상 반응시간의 허용 오차에 대한 최적화 연구 등이 수행되었다[14][15].

OTT 사업자는 다양한 유형의 고객들로부터 유동적인 클라우드 자원 요구량을 예측하여 미리 준비함으로써 오버 프로비저닝에 대한 기회비용과 언더 프로비저닝에 따른 지연시간 발생을 예방해야 한다. 이를 위해 자원 요구 변화에 따라 선제적으로 정확하고 신속하게 자원을 할당하는 기법 제시가 요구된다.

2-2 강화학습 기반 클라우드의 자원 할당 연구

OTT 서비스는 대규모의 컴퓨팅 및 네트워크 자원을 필요로 하는 미디어 서비스로서 사용자가 급격히 증가하는 시간대의 경우 트래픽이 폭증하여 서버에 과부하를 준다. OTT 사용자 및 콘텐츠의 유형은 매우 다양하여 트래픽이 증가하는 추세를 예측하기 어려우며, 이 경우 서버가 마비되는 현상이 발생할 수 있다. Hui Zhao 등의 연구에서는 VoD 사용자들의 비디오 시청 인기도 (Video Popularity)와 상영 시간 (Playing Time) 등을 포함하여 사용자들의 역동적인 행동 패턴을 분석한 후, 그 특성에 맞는 가상 스트리밍 미디어 서버 클러스터에 대한 리소스 할당 모델을 제안하였다[16].

강화학습은 어떤 환경에서 선택 가능한 행동들의 시행착오를 거쳐 주어지는 보상을 최대화하는 행동을 학습하는 방법이다[17]. 강화학습 환경에서는 행동하는 에이전트가 현재 상태를 인식하여 선택 가능한 행동 중 하나를 선택해 행동하고, 다음의 상태로 이동하면서 그 행동의 대가로 보상을 받게 된다. 강화학습 모델은 이러한 보상의 총합을 최대화하는 것이다. 주요 강화학습 기법인 Q-learning은 별도의 모델이 없어도 행동에 대한 예상 가치를 예측할 수 있으며, 마르코프 의사결정 프로세스(Markov Decision Process: MDP)에서 Q함수를 연구하여 최적의 정책을 학습한다[18].

근래에 VoD, OTT 플랫폼 등 미디어 서비스에서 다양한 유형의 사용자들과 역동적인 사용량 패턴에 적합한 자원 할당 기법을 개발하기 위하여 강화학습을 기반으로 한 클라우드 할당 관련 연구들이 진행되고 있다. 강화학습을 이용한 효율적인 클라우드 자원 자동 확장 연구에서는 SLA 위반을 최소화하기 위해 Q-learning, Learning Automata, Cost Aware의 접근법을 비교하였다. 그 결과 Q-learning 기법을

사용했을 때 SLA 위반을 크게 줄이고 시스템 안정성을 높일 수 있었다[19]. 또한, 무선 네트워크에서 Q-learning 기반의 강화학습 에이전트를 활용하여 데이터의 전송 속도 최적화하는 연구가 수행되었다[20]. 이러한 연구들을 통해 다양한 서비스 환경에서 자원 할당 등을 제어하는 데 있어서 강화학습 알고리즘이 효과적임을 확인할 수 있다.

본 연구에서 강화학습을 적용하기 위해서는 OTT 사용자들의 서비스 요청률(Demand)과 시간 변수를 고려해야 한다. 클라우드 환경에서 서비스 요청에 따라 강화학습 방식으로 CPU 및 네트워크 대역폭을 등을 할당하는 연구들이 수행되었으며[21][22][23], YouTube의 Open API 분석을 통한 글로벌 게임방송 스트리밍 현황 및 시청행태 데이터 속성 등이 기존 연구에서 분석되었다[24].

Ⅲ. 심층강화학습 기반 비용 효율적인 OTT 서비스 플랫폼

3-1 클라우드 기반 OTT 서비스 플랫폼 구조

클라우드 기반 OTT 서비스 플랫폼 구조는 그림 1과 같다. OTT 서비스 제공자는 상용 클라우드 컴퓨팅 서비스에 가입하여, 서비스 제공에 필요한 VM 인스턴스들을 할당받은 후, OTT 서비스 제공을 위해 필요한 OTT 서비스 플랫폼을 설치하여 운용한다.

OTT 서비스 플랫폼은 OTT 서비스 메뉴, 콘텐츠 목록 및 사용자 인터페이스를 담당하는 OTT 웹 서버 기능, 콘텐츠 저장을 위한 스토리지 기능, 미디어 스트리밍을 위한 스트리밍 기능, 콘텐츠 해상도 및 포맷 변환을 위한 트랜스코딩 기능, 그리고 클라우드 자원 요청 및 반납 등을 처리하는 심층강화학습 기반의 클라우드 자원 관리 기능 등으로 구성된다.

OTT 서비스 이용자는 태내에서 OTT 셋탑박스가 연결된 TV 또는 스마트폰 앱 등을 통해 OTT 서비스 플랫폼에 접속한 후에, 원하는 콘텐츠를 선택하여 시청할 수 있다. OTT 서비스 이용자와 OTT 서비스 플랫폼 간의 연결에는 유선 또는 무선 네트워크가 이용될 수 있을 것이다.

본 논문에서 제안하는 OTT 서비스 플랫폼 내의 심층강화학습 기반의 클라우드 자원 관리 기능은 과거 OTT 사용자들의 서비스 이용 데이터들을 바탕으로 접속량 변화 패턴 등을 학습한 후, OTT 서비스 요청의 증감 상황에서 요구되는 클라우드 자원량을 선제적으로 할당받거나, 불필요한 클라우드 자원을 사전에 반납함으로써 최적의 클라우드 자원을 유지하며 비용 효율적인 OTT 서비스를 제공한다.

예측된 자원량을 클라우드에 요구하면, 클라우드는 요청된 자원을 OTT 서비스 플랫폼을 위해 할당하며, OTT 서비스 플랫폼은 미디어 서비스를 요청하는 사용자들에게 할당받은 자원을 배분하여 스트리밍 서비스를 제공한다.

세부적으로 클라우드 자원 관리 기능은 LSTM 신경망과

Deep Q Network (DQN) 알고리즘에 기반한 강화학습을 이용하여 시간대별 자원 사용량을 예측하고, 클라우드에 자원 할당을 요청한다. 강화학습 시도 초기에는 예측된 자원량과 실제 자원량이 일치하지 않았지만, 시간이 지나면서 축적된 데이터와 Feedback 정보를 바탕으로 계속 학습을 시킨다면, 이용자의 수요 예측 성능을 향상할 수 있다.

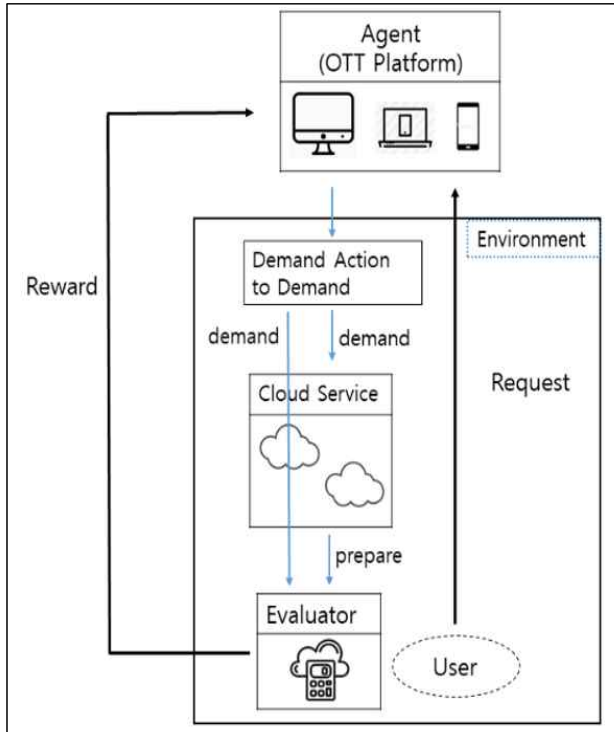


그림 1. 클라우드 기반 OTT 서비스 플랫폼 아키텍처
 Fig. 1. Cloud based OTT Service Platform Architecture

3-2 OTT 서비스 자원할당을 위한 심층강화학습 모델링

본 연구에서 클라우드 자원 관리 기능, 즉 에이전트는 인간의 개입 없이 OTT 서비스 워크로드에 따라 OTT 서비스 품질 저하를 최소화하면서, 최적의 클라우드 자원을 동적으로 할당하여 서비스 비용을 절감하는 것을 목표로 한다.

그림 1과 같은 시스템을 구축하기 위해 LSTM 신경망과 DQN 알고리즘에 기반한 강화학습 모델을 설계하였다. 강화학습은 MDP 요소들로 정의되고, 요소들로는 상태(S), 행동(A), 상태 전이 확률(P), 보상(R), 감가율(γ)이 있다. 상태는 OTT 자원의 효율적이고, 유연한 자원 할당을 위한 것으로, 제안모델에서 사용되는 정의는 표 1과 같다.

본 연구는 OTT 서비스 플랫폼 사용량과 관련되어 있으므로 시간에 따른 이용자의 자원 요청량과 시스템 사용량을 인식할 수 있는 정보를 포함하는 환경 상태를 정의하였다. 환경에서 강화학습 에이전트는 현재 상태를 인식하며 선택 가능한 행동 중에서 보상을 극대화하는 행동을 선택한다.

표 1. 강화학습을 위한 MDP 요소 정의

Table 1. Defining MDP Elements for Reinforcement Learning

	Definition
S	set of all possible states {VMs, membership(Basic, Premium), user id, contents id, resolution(HD, UHD), running time, time }
A	Set of all possible actions {ADD, REMOVE, HOLD}
P	probability of moving from states to next state when action a is taken
R	compensation to evaluate only current status and behavior {1, 0, -1}
γ	the value of a reward awarded

에이전트는 주변 상태에 따라 어떤 행동을 할지 판단을 내리는 주체이고, 에이전트가 속한 환경을 정의해야 한다. 상태는 에이전트의 행동에 따라 바뀌게 되고, 보상과 처벌을 받는다. 환경은 가상의 User, Demand Action, Cloud로 구성된다. 사용자의 접속량 (Demand)와 클라우드 환경에서 준비한 자원(Prepare)은 CPU, 메모리, 스토리지, 네트워크 대역폭 등을 포함한다.

OTT 플랫폼의 동작 시간은 $t=0, 1, 2 \dots$ 로 나뉘며, 하나의 시간 슬롯은 T_t 로 표시한다. S_t 는 가능한 모든 상태의 집합을 표시하며, s_t 는 현재 시점의 상황을 나타내는 값의 집합이다. 행동은 OTT 서비스 플랫폼의 클라우드 자원 관리 기능, 즉 에이전트가 취할 수 있는 선택지로 가능한 모든 행동의 집합을 $A(s_t)$ 라 하고, 특정 시간 t에서의 상태인 s_t 를 기반으로 선택한 행동을 a_t 라고 한다.

보상은 에이전트가 어떤 행동을 했을 때 따라오는 이득이다. 보상은 현재 상태와 행동에 대해서만 평가하는 즉각적인 값으로 r 로 표시한다. 누적 보상이 제일 높은 행동을 선택하여 장기적으로 가장 좋은 결과를 얻는 행동으로 이어져야 한다. 누적보상의 기댓값을 해당 상태 S 의 가치 함수로 정의하여 재귀적으로 나타낸 벨만 방정식(1)은 다음과 같다.

$$Q(s, a) = E_{s', r \sim p(\cdot | s, a)} [r + \gamma E_{a' \sim \pi} [Q(s', a')]] \tag{1}$$

위의 벨만 방정식에서 γ 은 감가율로 0에서 1사이의 값을 취하게 된다. 최적 정책 (Optimal policy)는 여러 정책 중 누적 보상의 기댓값이 가장 큰 정책이다. 방정식 (2)와 같이 다음 상태 s' 에서 가장 큰 Q 값을 이용하여 Q 함수를 갱신하며, 행동을 결정할 때 Q 값이 가장 높은 행동을 선택하면서 충분한 탐험을 수행한다. 탐험은 학습 초기에는 많이 진행하고 학습 후반으로 갈수록 적게하는 ϵ -greedy 기법을 사용한다.

$$Q^*(s, a) = \max_{\pi} Q_{\pi}(s, a) = \max_{\pi} E_{\pi} [G_t | S_t = s, A_t = a] \tag{2}$$

상태와 행동이 Q-table로 표현하기에 많은 경우의 수 (High dimension)를 갖는 경우에는 과장 추정 (Overestimation) 문제가 발생할 수 있다. 이를 방지하기 위해 LSTM 신경망을 결합한 Q-learning 알고리즘을 사용하였다. 인공신경망을 결합한 Q-learning 알고리즘인 DQN은 기존의 Q-table 대신 Network를 통해 상태공간을 표현하였고, Experience Replay 기법을 적용해 상태-행동 가치함수 (State-action Value Function)가 수렴(Convergence)되는 것을 가능하게 한다. DQN은 $Q(s, a, w(weight))$ 가 최적의 $Q(s, a)$ 와 근사화되도록 하는 것이 목표이며, Stochastic Gradient Descent를 통해 손실 함수 (Loss Function)을 최소화하는 방향으로 가중치 (Weight)를 학습한다.

$$loss\ function = (r + \gamma \cdot \max_a Q(s', a', w) - Q(s, a, w))^2 \quad (3)$$

이에 따른 전체적인 프로세스는 아래의 표 2와 같다.

표 2. OTT 자원할당을 위한 LSTM 기반 DQN 알고리즘
Table 2. LSTM-based DQN algorithm for OTT resource allocation

Algorithm 1: Reinforcement Learning Algorithm (DQN)
Initialize replay memory D to capacity N
Initialize action-value function Q with random weights
For episode = 1, M do
For t = 1, T do
Apply LSTM classifier
input x is the OTT data, output is the action probability ϵ
With the probability ϵ select a action a_t
Execute action a_t and observe r_t
Set $s_{t+1} = s_t$
Store transition (x_t, a_t, r_t, x_{t+1}) in D
Sample random minibatch of transition (x_j, a_j, r_j, x_{j+1}) from D
Set $y_j = \begin{cases} r_j & \text{if episode terminate at step } j+1 \\ r_j + \gamma \max_a Q(x_{j+1}, a; \theta) & \text{otherwise} \end{cases}$
Perform a stochastic gradient descent
End for
End for

Mini-batch 형식으로 데이터를 저장하기 위해 replay memory를 생성하고 Q-learning을 위해 Q-value를 초기화 시킨다. Input data로 OTT 데이터가 들어오며, 해당 데이터는 유저의 정보와 유저가 시청하는 콘텐츠 정보 및 요구되는 자원량 정보를 담고 있다.

이후 LSTM 신경망을 사용하여 행동을 결정하며 결정된 행동을 바탕으로 Agent가 보유하고 있는 자원량과 요청되는 자원량을 비교하여 Provisioning의 성공 여부를 판단하고 보상을 부여한다. Under-provisioning이 발생하거나 Over-provisioning이 발생하는 경우 패널티를 부여하고, Optimal-provisioning인 경우에는 보상을 부여한다.

IV. 시뮬레이션 및 성능 분석

4-1 심층강화학습 기반 OTT 시뮬레이터 구현

강화학습 관련 RLtrader 오픈소스[25]를 수정 및 확장하여 OTT 데이터를 학습하고 클라우드 자원을 동적으로 할당 및 해제하는 OTT 서비스 플랫폼 시뮬레이터를 구현하였다.

구현된 심층강화학습 기반 OTT 자원할당 시뮬레이터에서 가상의 사용자들이 OTT 서비스 플랫폼에 접속하여 OTT 콘텐츠를 시청한다. 사용자들의 OTT 콘텐츠 시청에 따라서 요구되는 자원량은 시간대별로 달라진다. 시간대는 0에서 24시간 범위에서 1시간 단위로 설정한다. 사용자들의 OTT 콘텐츠 시청으로 인해 요구하는 자원량을 Request로 정의한다. OTT 서비스 플랫폼의 클라우드 자원 할당 기능, 즉 심층강화학습 에이전트는 콘텐츠 시청으로 인해 요구되는 자원량 Request에 따라 3.2절에 기술된 심층강화학습 알고리즘을 바탕으로 행동을 결정한다. 행동의 종류는 ADD, REMOVE, HOLD이다.

사용자들의 Request에 따라 에이전트에서 계산된 자원 요구량을 바탕으로 클라우드는 자원을 할당해주게 된다. 이렇게 할당된 클라우드 자원을 Prepare로 정의한다. Prepare에 대한 평가를 위해 환경 (Environment) 기능을 구현하였으며, 해당 Prepare에 따라서 optimal-/under-/over-provisioning, 즉 성공/실패/잉여자원 세 가지 경우가 결정되고, 그에 따라 환경이 보상과 변화를 부여한다.

에이전트가 이러한 환경과 상호작용하면서 사용자의 OTT 서비스 요청에 따라 클라우드 자원을 적응적으로 할당하기 위해 에이전트는 LSTM 신경망과 심층강화학습 모듈을 포함한다. 시뮬레이터 구현과 실행에 이용된 서버 환경과 시뮬레이션 파라미터 설정은 각각 아래 표 3, 4와 같다.

표 3. OTT 서비스 플랫폼 시뮬레이터 개발 환경
Table 3. OTT Service Platform Simulator Development Environment

Type	Environment
Operating System	Windows 10
CPU	Intel(R) Core(TM) i7-10875H 2.30 GHz
RAM	16.0GB
GPU	NVIDIA GeForce RTX 2060 Intel(R) UHD Graphics
Library	Tensorflow 2.6.0 Keras 2.6.0
Programming Language	Python 3.8.3

표 4. 시뮬레이션 파라미터들

Table 4. Simulation Parameters

Simulation parameters	Values
Neural network	LSTM
Activation argument	Sigmoid, ReLU
Regression loss	MSE
Epochs	500
learning rate	0.001
discount factor	0.95
Start exploration	0.5

본 논문에서 구현한 OTT 서비스 플랫폼 시뮬레이터는 OTT 서비스 요청 데이터를 입력받아 동작한다. 상용 OTT 서비스에서 평일에 OTT 서비스를 이용하는 사용자의 패턴을 기반으로 데이터를 무작위로 생성하여 강화학습에 적용하였다. OTT 서비스를 제공하기 위한 클라우드 자원은 네트워크 대역폭을 기준으로 적용하였으며, OTT 서비스 요청 개별 건별로 클라우드 자원이 적응적으로 할당되는 것을 가정하였다.

그림 2는 이 시뮬레이션에 적용된 하루 동안의 OTT 서비스 요청 패턴을 보여주며, X 축은 시간, Y 축은 사용자들로부터의 OTT 서비스 요청을 의미한다. 일반적으로 OTT 시청률은 0시~3시에 증가하고, 낮에는 감소하다가 18시부터 OTT 시청률이 다시 증가한다.

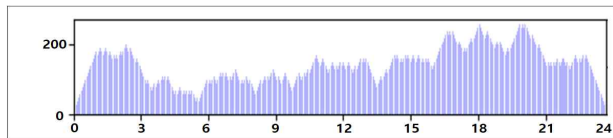


그림 2. OTT 서비스 요청 데이터
Fig. 2. OTT Service Request Data

4-2 성능 분석

3장에서 기술한 바와 같이 제안하는 OTT 플랫폼은 DQN 알고리즘을 기반으로 사용자의 OTT 서비스 요청에 따라 클라우드 자원을 최적으로 할당 및 해제함으로써 고품질의 OTT 서비스를 저렴한 비용으로 제공하는 것을 목표로 한다.

앞서 설명된 OTT 서비스 플랫폼 시뮬레이터에 대해 Sigmoid와 ReLU(Rectified Linear Unit) 활성화 함수를 각각 적용하고, epoch는 500으로 설정하여 에이전트에서 LSTM 기반 DQN 알고리즘을 통한 클라우드 자원 관리 기능 동작을 분석하였다. 분석 시 과적합(Overfitting)을 방지하기 위해 학습이 진행되는 동안 성능의 변화가 없거나 과적합 발생 이후 30 epoch 동안 성능이 향상되지 않는 경우 학습을 조기종료 하여 모델을 구축하였다.

에이전트가 학습을 시작하는 초기 상태에서 에이전트는 자

원 할당 및 반납 요청을 위한 임의의 작업을 수행한다. 자원 요청 및 반납의 결과에 따라서, optimal-/under-/over-provisioning 여부를 판단하고 긍정 또는 부정적인 보상이 주어진다. 에이전트는 환경에서 받은 상태, 보상, 액션 데이터를 입력하여 학습을 수행하고 그에 따라 클라우드 자원 할당을 제어하는 행동을 갱신한다.

Sigmoid 함수를 적용한 모델인 그림 3은 에이전트의 강화 학습 과정을 반복하면서 클라우드 자원을 할당하고 해제한 결과를 보여준다. X축은 시간, Y축은 할당된 OTT 클라우드 자원을 의미한다. 초기 epoch에서는 에이전트에서 학습이 수행되기 전에 사용자의 OTT 서비스 요청에 맞지 않는 클라우드 자원할당 상태를 볼 수 있다. epoch의 수가 증가할수록 에이전트의 학습이 향상되고 오류가 줄어든다. epoch가 200일 때, OTT 서비스 요청과 유사한 형태로 클라우드 자원이 할당되는 것을 확인할 수 있다. 그림 3은 epoch 횟수가 충분히 증가하여 500일 때, 서비스 요청에 가장 근접한 형태로 클라우드 자원이 동적으로 할당 및 해제되는 것을 보여준다.

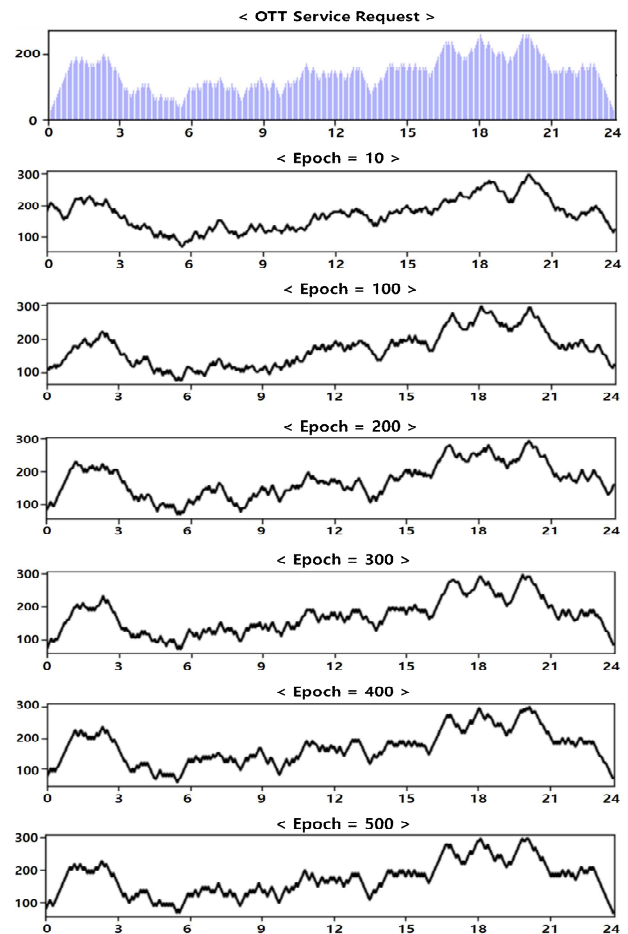


그림 3. Sigmoid를 적용한 심층강화학습 기반 OTT 클라우드 자원할당 성능
Fig. 3. OTT Cloud Resource Allocation Performance based on Sigmoid Applied Deep Reinforcement Learning

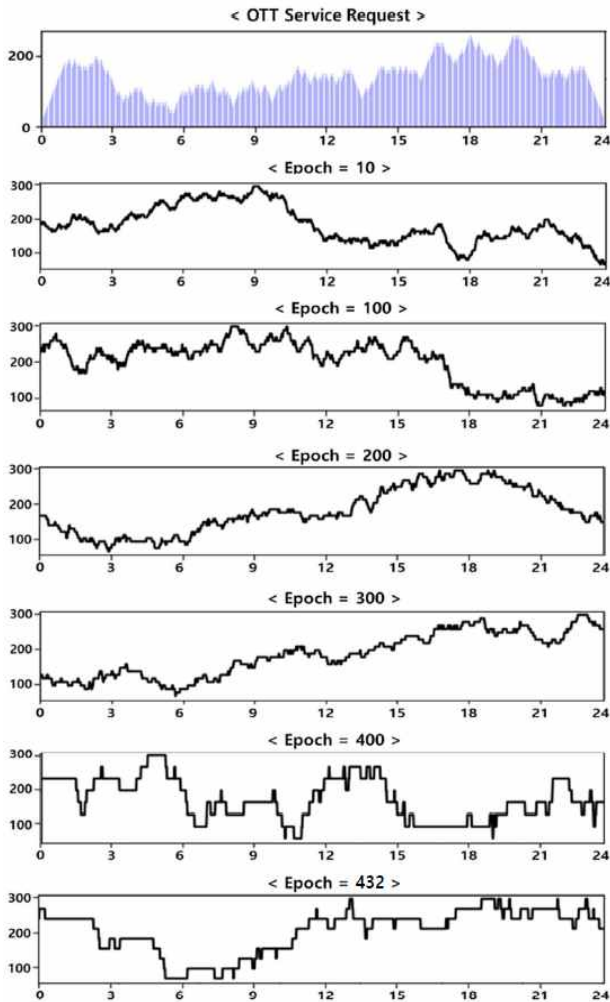


그림 4. ReLU를 적용한 심층강화학습 기반 OTT 클라우드 자원할당 성능
 Fig. 4. OTT Cloud Resource Allocation Performance based on Sigmoid Applied Deep Reinforcement Learning

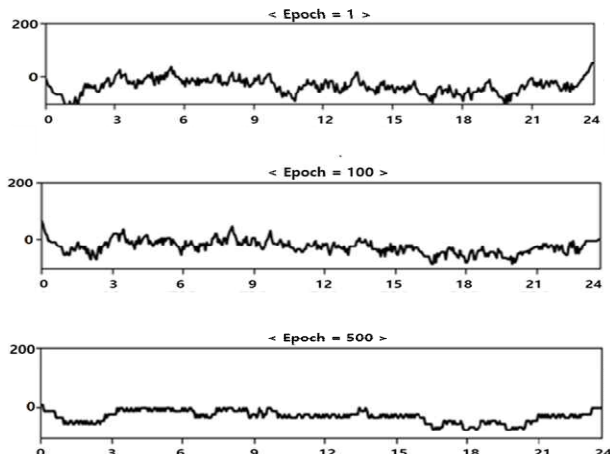


그림 5. OTT 서비스를 위해 확보된 클라우드 자원의 양
 Fig. 5. OTT Cloud Resource Allocation Performance based on Deep Reinforcement Learning

ReLU 함수를 적용한 모델은 그림 4와 같으며, epoch가 증가함에 따라서 비교적 요청량에 근접한 자원을 할당하고 해제하는 동작을 수행하지만, Sigmoid 함수를 적용한 모델이 ReLU 함수에 비해 상대적으로 요청량에 근접한 자원을 동적으로 할당하는 것으로 확인되었다.

그림 5는 OTT 서비스 요청에 대해 제안된 LSTM 기반 DQN 방식의 클라우드 자원 관리를 통해 확보된 클라우드 자원의 양을 보여준다. 본 그래프에서 X축은 시간, Y축은 확보된 클라우드 자원의 양을 의미한다. 학습 초기 단계(epoch=1)에서는 임의의 자원할당 및 해제에 대한 시도가 이뤄지므로 OTT 서비스 요청 대비 확보된 클라우드 자원이 부족하여 OTT 서비스 요청을 수락할 수 없는 Under-provisioning, 또는 과도한 클라우드 자원이 할당되어 자원을 낭비하게 되는 Over-provisioning 상황이 발생하는 것을 확인할 수 있다. 또한, Optimal-provisioning을 의미하는 0 값을 기준으로 상하로 100에 가까운 변동이 빈번하여 자원의 낭비와 OTT 서비스 품질 저하가 크게 발생하는 것을 알 수 있다.

그러나, epoch의 수가 증가할수록 에이전트의 학습이 향상되고 오류가 감소하므로, 그림 5의 epoch가 500일 때와 같이, 확보된 자원이 Optimal-provisioning을 의미하는 0에 근접하게 수렴하고, 그 변동 폭이 좁아져서 안정적으로 자원 할당 및 해제가 수행됨을 확인할 수 있다. 이를 통해, 제안 방식은 OTT 서비스 플랫폼의 클라우드 자원 관리 측면에서 OTT 서비스 품질 저하를 최소화하면서 OTT 서비스를 위한 클라우드 자원 낭비를 줄여서 가격 경쟁력 있는 OTT 서비스 지원할 수 있을 것이다.

V. 결 론

본 논문에서는 현재 빠르게 확산되고 있는 클라우드 기반의 OTT 서비스 플랫폼에서 사용자에게 제공되는 OTT 미디어 서비스의 품질을 유지하면서도 요구되는 클라우드 자원을 최소화하여 비용 효율적인 OTT 서비스를 제공할 수 있는 심층 강화학습 기반의 OTT 클라우드 자원 관리 방법을 제안하였다. 시뮬레이션을 통해 LSTM 기반의 DQN 알고리즘을 이용하는 제안 모델이 과거의 OTT 시청 데이터를 학습하여 OTT 서비스 요청이 급변하는 상황에서 OTT 서비스 요청을 수용할 수 있는 충분한 양의 클라우드 자원을 선제적으로 할당하거나 불필요한 클라우드 자원을 반납하는 방식으로 OTT 서비스를 위한 클라우드 자원을 효과적으로 관리함을 입증하였다.

마지막으로 향후 연구를 위한 몇 가지 방향을 제안한다. 첫째로 현재 연구는 클라우드 인스턴스 임대 비용과 용량, 시간 대별 사용량을 고려한 성능분석 및 검증이 요구된다. 둘째, 이 연구에서는 동일한 등급의 OTT 서비스 이용자들만을 대상으로 하였으나, Netflix 등의 상용 OTT 서비스와 같이 차별화

된 서비스 등급을 제공하는 상황에서의 비용 효율적인 자원 할당 방법에 관한 추가적인 연구가 필요할 것이다.

감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(NRF-2018R1A2B2003774) 및 정보통신기획평가원의 지원(2021-0-02068, 인공지능 혁신 허브 연구 개발)을 받아 수행된 연구임.

참고문헌

- [1] H. C. Yang, "Contents Industry Trend of USA," *Korea Creative Content Agency*, Vol. 21, November 2019.
- [2] K. H. Lyum, K. H. Kim and K. S. Kim, "AWS Cloud based Failover Load Balancing Mechanism and Availability Assessment," *KIISE Transactions on Computing Practices*, Vol. 24, No. 5, pp. 211-219, May 2018.
<https://doi.org/10.5626/KTCP.2018.24.5.2>
- [3] C. T. Yang, S. T. Chen, Y. W. Chan and Y. C. Shen, "On construction of a cloud storage system with heterogeneous software-defined storage technologies," *Human-centric Computing and Information Sciences*, Vol. 9, No. 12, pp. 1-17, April 2019.
<https://doi.org/10.1186/s13673-019-0173-x>
- [4] W. Liu, P. Wang, Y. Meng, C. Zhao and Z. Zhang, "Cloud spot instance price prediction using kNN regression," *Human-centric Computing and Information Sciences*, Vol. 10, No. 34, pp. 1-14, August 2020.
<https://doi.org/10.1186/s13673-020-00239-5>
- [5] M. S. Park, Y. S. Lee, S. B. Oh, M. K. Jo, H. R. Jang, Y. I. Yoon and T. W. Um, "A Study on Cloud Resource Allocation for Efficient Online Video Service Based on Reinforcement Learning," *Proceedings of The 12th International Conference on Computer Science and its Applications(CSA 2020)*, pp. 14-16, December 2020.
- [6] R. Bouabdallah, S. Lajmi and K. Ghedira, "Use of Reactive and Proactive Elasticity to Adjust Resources Provisioning in the Cloud Provider," *IEEE 18th International Conference on High Performance Computing and Communications*, pp. 1155-1162, December 2016.
<https://doi.org/10.1109/HPCC-SmartCity-DSS.2016.0162>
- [7] W. H. Liao, S. C. Kuai and Y. R. Leau, "Auto-scaling strategy for amazon web services in cloud computing," *In 2015 IEEE International Conference on Smart City/SocialCom/SustainCom(SmartCity)*, pp. 1059-1064, December 2015.
<https://doi.org/10.1109/SmartCity.2015.209>
- [8] D. K. Kim, S. H. Kim and C. H. Youn, "Cost Efficient Virtual Machine Brokering in Cloud Computing," *KIPS Transactions on Computer and Communication Systems*, Vol. 3, No. 7, pp. 219-230, July 2014.
<http://dx.doi.org/10.3745/KTCCS.2014.3.7.219>
- [9] Y. Zhang, J. Yan and H. Guan, "Intelligent Cloud Resource Management with Deep Reinforcement Learning," *IEEE Cloud Computing*, Vol. 4, No. 6, pp. 60-69, December 2017.
<http://dx.doi.org/10.1109/MCC.2018.1081063>
- [10] S. Horovitz and Y. Arian, "Efficient Cloud Auto-Scaling with SLA objective using Q-Learning," *IEEE 6th International Conference on Future Internet of Things and Cloud*, pp. 85-92, August 2018.
<http://dx.doi.org/10.1109/FiCloud.2018.00020>
- [11] K. W. Au, "Optimizing Auto-scaling Virtual Machines for a Cloud-based VOD Data Center," Master Degree Dissertation, *Graduate School, Hong Kong University*, 2016.
- [12] N. G. Loung and C. W. Ping, "Resource Management in Cloud Networking Using Economic Analysis and Pricing Models: A Survey," *IEEE Communications Surveys & Tutorials*, Vol. 19, No.2, pp. 954-1001, January 2017.
<http://dx.doi.org/10.1109/COMST.2017.2647981>
- [13] Y. H. Choi, Y. J. Lin and J. S. Park, "Reinforcement Learning Approach for Resource Allocation in Cloud Computing," *The Journal of the KICS*, Vol. 40, No. 4, pp. 653-658, April 2015.
<https://doi.org/10.7840/kics.2015.40.4.653>
- [14] M. D. Assuncao, H. Cardonha, Carlos, A. S. Netto, Marco and L. G. Cunha, Renato, "Impact of User Patience on Auto-Scaling Resource Capacity for Cloud Services," *Future Generation Computer Systems*, Vol. 55, pp. 41-50, September 2015.
<https://doi.org/10.1016/j.future.2015.09.001>
- [15] T. Ruether, "2019 Video Streaming Latency Report," *Wowza Media Systems*. Available:
<https://www.wowza.com/wp-content/uploads/Streaming-Video-Latency-Report-Interactive-2020.pdf>, January 2021.
- [16] H. Zhao, J. Wang, F. Liu, Q. Wang, N. Luo and W. Zhang, "Resource Allocation for Virtual Streaming Media Server Cluster in Cloud-based Multi-version VoD," *IEEE 22nd International Conference on Computer Supported Cooperative Work in Design*. pp. 313-318, May 2018.
<http://dx.doi.org/10.1109/CSCWD.2018.8465244>
- [17] S. C. Woo and Y. S. Sung, "Dynamic Action Space Handling Method for Reinforcement Learning Models," *Journal of Information Processing Systems*, Vol. 16, No.

- 5, pp. 1223-1230, October 2020.
<http://dx.doi.org/10.3745/JIPS.02.0146>
- [18] J. S. Park and J. H. Park, "Enhanced Machine Learning Algorithms: Deep Learning, Reinforcement Learning, and Q-Learning," *Journal of Information Processing Systems*, Vol. 16, No. 5, pp. 1001-1007, October 2020.
<http://dx.doi.org/10.3745/JIPS.02.0139>
- [19] B. Asgari, M. G. Arani and S. Jabbehdari, "An Efficient Approach for Resource Auto-Scaling in Cloud Environments," *International Journal of Electrical and Computer Engineering*, Vol. 6, No. 5, pp. 2415-2424, October 2016. <http://dx.doi.org/10.11591/ijece.v6i5.10639>
- [20] S. H. Cho, "Rate adaptation with Q-learning in CSMA/CA wireless networks," *Journal of Information Processing Systems*, Vol. 16, No. 5, pp. 1048-1063, October 2020.
<https://doi.org/10.3745/JIPS.03.0148>
- [21] M. K. Lee, "Reinforcement Learning based Service Deployment for Optimal Resource Allocation in 5G Network Slicing," Master Degree Dissertation, *Graduate School, Kyunghee University*, 2019.
- [22] K. H. Kim, I. Y. Jung, H. M. Kim, Y. J. Kim and C. S. Jeong, "SLA-based Auto-Scaling Engine for Cloud based Collaboration Platform," 2014 *Korea Electronic Engineering Association Summer General Academic Conference*, Vol. 37, No. 1, pp. 1518-1521, June 2014.
- [23] Y. W. Kim and J. C. No, "Resource Allocation based on Virtual Machine Resource Requirement Estimation in Cloud Computing Environment," *Journal of the Institute of Electronics and Information Engineers*, Vol. 54, No. 11, pp. 113-121, November 2017.
<https://doi.org/10.5573/ieie.2017.54.11.113>
- [24] M. J. Choe, H. R. Yong and H. S. Hwang, "A Study of Analyzing Live Streaming OTT Service Data : Focused on Youtube Game Broadcasting," *Journal of Korea Game Society*, Vol. 16, No. 2, pp. 61-73, April 2016.
<https://doi.org/10.7583/JKGS.2016.16.2.61>
- [25] QuantyLab, "RLtrader open source," Available: <https://github.com/quantyLab>



이예슬(Ye-Seul Lee)

2015년 : 숙명여자대학교 대학원 (무용학석사)

2019년 : 숙명여자대학교 대학원 (무용학박사)

2020년~현 재: 숙명여자대학교 빅데이터분석융합학과 석사과정

※ 관심분야 : Cloud Computing, Kinect Method, Motion Capture



장혜림(Hye-Rim Jang)

2019년 : 숙명여자대학교 (정치학사, 경제학사)

2019년~현 재: 숙명여자대학교 빅데이터분석융합학과 석사과정

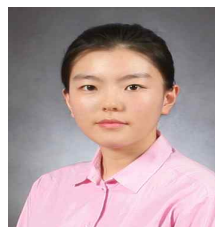
※ 관심분야 : Cloud Computing, Machine Learning, Deep Learning



오수빈(Soo-Been Oh)

2020년 : 숙명여자대학교 사회심리학 (문학사)

2020년~현 재: 숙명여자대학교 빅데이터분석융합학과 석사과정
※ 관심분야 : Cloud Computing, Deep Learning, Prediction



이연수(Yeon-Su Lee)

2018년 : 숙명여자대학교 (법학사)

2020년~현 재: 숙명여자대학교 빅데이터분석융합학과 석사과정
※ 관심분야 : Cloud Computing, Machine Learning, Text Mining, Personal Information



윤용익(Yong-Ik Yoon)

1983년 : 동국대학교 통계학과 (공학사)
1985년 : 한국과학기술원 전산학과 (공학석사)
1994년 : 한국과학기술원 전산학과 (공학박사)

1985년~1997년: 한국전자통신연구원 책임연구원
1997년~현 재: 숙명여자대학교 IT공학전공 교수 및 빅데이터분석융합학과 주임교수
관심분야 : 인공지능 서비스, 빅데이터 서비스, 지능형 IoT, 미들웨어, 디지털트윈



엄태원(Tai-Won Um)

1994년 : 홍익대학교 전자전기제어공학과 (공학사)
2000년 : 한국과학기술원 정보통신대학 (공학석사)
2006년 : 한국과학기술원 정보통신대학 (공학박사)

2006년~2017년: 한국전자통신연구원 책임연구원
2017년~2020년: 조선대학교 정보통신공학부 부교수
2020년~2022년: 덕성여자대학교 사이버보안전공 조교수
2022년~현 재: 전남대학교 데이터사이언스대학원 부교수
관심분야 : 지능형 미디어, 자율형 IoT, ICT 트러스트, 사이버물리시스템(CPS)