

5G 스몰셀 네트워크에서 기계 학습을 이용한 무선 환경 학습 및 오프로딩 기법

안상현*, 김종원*, 전우선*, 소정민^o

An Offloading Scheme through Wireless Environment Learning for 5G Small Cell Network

Sanghyeon An*, Jongwon Kim*, Woosun Chon*, Jungmin So^o

요약

밀리미터파의 경우 주파수 대역이 높기 때문에 감쇠가 심해 좁은 영역에서만 통신이 가능하다. 스몰 셀 기술은 무선 통신 기기가 증가함에 따라 여러 개의 소형 기지국을 작은 영역으로 나누어 설치하는 기술인데, 스몰 셀 기술은 5G 네트워크에서 밀리미터파 대역 지원에 필수적인 기술로 알려져 있다. 그러나 스몰 셀 기술은 통신 영역이 좁고 다수의 요청을 한 번에 처리하기 어렵다. 특히, 스몰 셀에서 통신이 폭발적으로 증가할 경우 인근 스몰 셀 기지국이나 매크로 셀 기지국에서 처리해줘야 한다. 본 논문에서는 셀룰러 네트워크 환경에서 기계 학습을 이용한 각 셀 간 오프로딩 기법을 제안한다.

Key Words : Cellular Network, Machine Learning, Offloading, Small Cell Technology, 5G Network

ABSTRACT

In the case of millimeter wave, because the frequency band is high, the attenuation is severe so communication is possible only in a small area. Small cell technology is to install several small base stations by dividing them into small areas due to the increase in wireless communication devices. This small cell technology is known as an essential technology for supporting the millimeter wave band in a 5G network. However, small cells have a narrow range of communication areas and are difficult to handle a large number of requests. In particular, when communication explosively increases in a small cell, there is a problem that a offloading needs to be made to a nearby small cell base station or a macro cell base station. In this paper, to optimize the offloading, we propose the offloading algorithm through machine learning among each cell.

1. 서론

5G 통신 기술의 발전으로 무선 통신의 수는 기하급수적으로 증가하였고 높은 속도의 데이터 속도를

요구하기 시작하였다. 많은 데이터 요청을 처리하고 데이터 속도를 증가시키기 위해 다양한 기술들이 개발되었다. 특히, 5G는 현재 상용화된 Sub-6GHz 대역 뿐 아니라 매우 높은 주파수 대역인 28GHz~300GHz

* 이 논문은 2019년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2019R1A2C1005881).

• First Author : Sogang University Department of Computer Science and Engineering, ansh941@sogang.ac.kr, 학생(석사), 학생회원

◦ Corresponding Author : Sogang University Department of Computer Science and Engineering, jso1@sogang.ac.kr, 교수, 정회원

* Sogang University Department of Computer Science and Engineering, kjo26619@sogang.ac.kr, 학생(석사); sunnyjun95@sogang.ac.kr, 학생(석사),

논문번호 : 202008-195-B-RN, Received August 12, 2020; Revised September 14, 2020; Accepted October 26, 2020

인 밀리미터파 대역을 사용하는 두 가지 기지국으로 나뉘어 졌다.^[1]

5G에서 사용되는 전자기파는 높은 주파수 대역을 사용하고 있어 파장이 짧기 때문에 비교적 회절 현상이 약해 신호를 멀리 보내기는 힘들지만 주파수가 증가할수록 높은 데이터 속도와 직진성을 가져 특정 방향으로 데이터를 보내는데 유리하다. 하지만, 장애물에 따른 감쇠 현상으로 인해 전파가 통과하지 못하는 경향이 있어 밀리미터파 대역은 작은 영역의 통신만 가능하다. 하지만, 5G 네트워크의 발전에서 밀리미터파 대역의 지원은 반드시 필요하기 때문에 밀리미터파 대역을 사용하는 기지국은 스몰 셀 기술과 함께 사용된다.

스몰 셀 기술은 점차 증가하는 무선 매체 사용자 및 장애물에 따라 기존의 넓은 통신 영역을 갖던 하나의 셀을 작은 영역으로 나누는 기술이다. 하나의 매크로 셀에는 여러 개의 스몰 셀이 존재할 수 있으며 사용자는 매크로 셀 기지국이나 스몰 셀 기지국(SBS: Small cell Base Station)을 선택하여 통신할 수 있다. 이 기술은 본래 많은 양의 통신이 발생하는 구역에 두어 데이터를 오프로딩 하기 위해 등장하였다. 또한, 5G 네트워크가 도입되면서 매크로 셀 기지국(MBS: Macro cell Base Station)은 Sub-6GHz 대역을 사용하고 스몰 셀 기지국은 밀리미터파 대역을 사용하는 기본적인 5G 네트워크 구조가 되었다. 하지만, 스몰 셀 기지국은 비교적 낮은 파워와 설치 및 유지 비용 등의 이유로 한 번에 많은 양의 요청을 처리할 수는 없다. 높은 데이터 속도를 요구하는 5G 사용자는 밀리미터파 대역을 이용하면서 가까운 스몰 셀 기지국에게 요청을 보내지만 스몰 셀이 전부 처리할 수 없는 상태가 발생할 수 있다.

이러한 문제점을 해결하기 위해 기존의 방법에서는 스몰 셀 기지국에 많은 양의 요청이 들어오면 요청 사용자와 매크로 셀 기지국과의 거리를 고려하여 최적의 기지국에 할당하는 알고리즘을 제안하였다.^[2]

하지만, 거리만 고려할 경우 사용자마다의 지연 민감, 자원 요청과 같은 다양한 요구사항이 무시되어 다양한 QoS를 보장해야 하는 5G 네트워크에서 효율적이지 못하다는 문제점이 생긴다. 이러한 문제점을 해결하기 위해 본 논문에서는 기계 학습을 활용하여 통신 환경을 학습하고 스몰 셀 기지국에서 다양한 요구사항을 갖는 사용자 중 매크로 셀 기지국에서 처리하는 것이 더 효율적인 사용자의 경우에는 오프로딩할 수 있도록 한다. 제안하는 오프로딩 알고리즘은 다음과 같다.

먼저, 매크로 셀 기지국에서 주기적으로 스몰 셀을 관찰하면서 스몰 셀 내부의 사용자들의 다양한 사용자 정보를 수집한다.

매크로 셀 기지국에서 수집한 다양한 정보로는 사용자의 위치, 매크로 셀 기지국과의 거리, 요청 지연 정도, 요청 패킷 크기, 통신 성공 여부 등과 같은 정보이며, 이 다양한 정보를 토대로 기계 학습을 활용하여 사용자와의 채널 상태를 학습하고 매크로 셀 기지국과의 통신에 있어 가장 최적의 사용자를 선택하여 데이터 오프로딩 여부를 결정한다.

마지막으로, 본 논문에서 제안한 다양한 정보를 모두 학습하여 데이터 오프로딩 여부를 결정할 기법과 기존의 거리만 가지고 데이터 오프로딩 여부를 결정할 기법을 비교하여 본 논문에서 제안한 기법이 오프로딩을 통해 효율적으로 자원을 활용할 수 있었음을 보여준다.

II. 본 론

본 장에서는 우리가 제안하는 네트워크의 구조와 사용자 분류를 위해 사용한 데이터 마이닝 기법과 기준을 소개 후 이를 기반으로 데이터 오프로딩 여부를 결정하는 기준을 소개한다.

2.1 네트워크 구조

본 논문에서 고려하는 네트워크 구조는 그림 1과 같다. 하나의 매크로 셀 기지국(MBS)과 M개의 스몰 셀 기지국(SBS), 그리고 N명의 사용자가 각자 임의로 배치되어 있다.

매크로 셀 기지국에서는 주기적으로 스몰 셀 기지국으로부터 사용자들의 위치, 요청 패턴, 요청 패킷 크기, 요청 지연, 요구 처리량(Throughput) 등 다양한 정보를 수집한다. 수집한 정보를 토대로 매크로 셀 기지국에서는 기계 학습을 통해 사용자와 매크로 셀 기

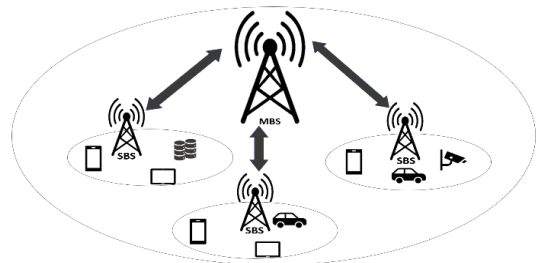


그림 1. 네트워크 구조
Fig. 1. Network Structure

지국과의 통신 적합성 기준으로 데이터 오프로딩 여부를 결정한다.

네트워크의 무선 채널은 경로 손실(Path Loss)을 고려한다. 특히, 몇몇의 장소에는 장애물이 있어서 각 기지국과 사용자 간의 경로 손실이 더욱 심해진다. 경로 손실이 심하면 사용자는 제대로 패킷을 받지 못하고 재요청이 발생하여 전체 네트워크 성능이 떨어질 수 있다.

장애물이 없는 상태를 LOS(Line Of Sight), 장애물이 있는 상태를 NLOS(Non Line Of Sight)라고 하며 각각의 경로 손실은 다음과 같이 정의한다.^[3]

$$PL_{LOS} = 28.0 + 22\log_{10}(d_{3D}) + 20\log_{10}(f_c) \quad (1)$$

$$PL_{NLOS} = 13.54 + 39.08\log_{10}(d_{3D}) + 20\log_{10}(f_c) - 0.6(h_{UT} - 1.5) \quad (2)$$

여기서, d_{3D} 는 3차원 상의 기지국과 사용자간의 거리이다. f_c 는 각 기지국이 사용하는 중심 주파수이다. h_{UT} 는 사용자의 높이이다.

각 기지국들은 정확한 네트워크의 상황과 특정 위치에 장애물이 있는지 알 수 없다. 네트워크 상황과 장애물 유무 등을 파악하기 위해서는 이전의 통신 결과를 통해서 학습해야 한다. 수집한 데이터를 기계 학습을 이용하여 학습한 후 사용자 요청이 들어오면 어느 기지국이 처리하는 것이 가장 좋은지 결정한다. 다음 절에서는 다양한 QoS를 가진 사용자들의 통신 결과를 학습하는 기계 학습 기법을 소개한다.

2.2 기계 학습 및 학습 모델

본 절에서는 매크로 셀 기지국에서 데이터 오프로딩을 결정하기 위해 사용하는 기술인 기계 학습에 대하여 설명하고 사용한 기법을 소개한다.

기계학습이란 인공지능의 한 분야이며 컴퓨터가 처리하는 알고리즘을 만드는 것이 아닌 컴퓨터 스스로가 학습하여 데이터를 처리할 수 있게 학습하는 알고리즘을 만드는 이론이다. 기계 학습은 지도(Supervised) 학습과 비지도(Unsupervised) 학습으로 나뉜다.

지도 학습은 입력 데이터와 쌍으로 이루어진 정답들이 있어 학습 모델에게 정답을 보여주면서 모델이 입력 데이터에 대해 올바르게 판단할 수 있도록 알려주는 학습 방법이다. 지도 학습으로 풀어낼 수 있는 문제 중 분류 문제는 입력 데이터가 어떤 그룹에 속하는지 분류하는 문제이다. 본 논문에서는 데이터 오프로딩 여부를 지도 학습의 분류 문제로 정의하여 풀어

냈다.

2.2.1 의사 결정 트리(Decision Tree)

의사 결정 트리는 트리 기반 학습 모델로 예/아니오 질문을 이어가며 학습하는 모델이다.^[4] 그림 2는 의사 결정 트리의 구조 및 예시를 나타내고 있는데, 이와 같이 특정 기준에 따라 데이터를 구분하는 모델을 결정 트리 모델이라고 한다. 결정 트리 모델에서는 어떤 기준으로 노드를 놓아야 하는지와 어떤 노드를 루트에 놓아야 하는지가 중요한데 이것들을 효율적으로 배치하기 위해 정보 획득량(Information gain)을 사용한다.

하지만 의사 결정 트리의 경우 트레이닝 데이터에 약간의 차이만 발생해도 트리의 모양이 많이 달라질 수 있다. 이것은 모델이 입력 데이터에 매우 민감하다는 것을 의미한다. 따라서 과적합(Overfitting)에 매우 취약하다.

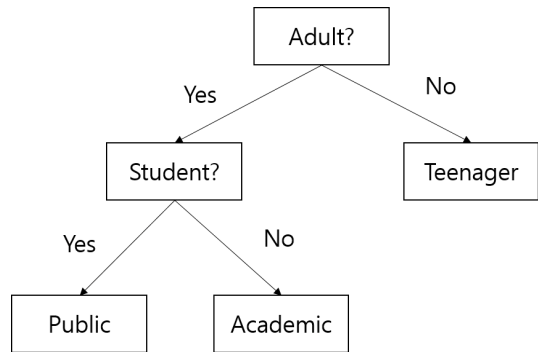


그림 2. 의사 결정 트리의 구조 및 예시
Fig. 2. Structure and example of Decision Tree

2.2.2 Random Forest

Random Forest는 대표적인 Bagging 앙상블 기반 모델로, 하나로 과적합을 방지하기 위해 최적의 기준 변수를 랜덤 선택하는 Breiman이 제안한 머신러닝 기법이다.^[5] Random Forest 모델은 먼저 데이터셋을 서로 교집합이 존재하는 n개의 서브셋으로 나누어 n개의 결정 트리 모델에 각각 학습시킨 후, 과반수 투표(Majority voting) 알고리즘을 이용해 최종적인 답을 내는 알고리즘이다. 단일 의사 결정 트리에 비해 결측치(Missing value)를 다루기 쉬우며 대용량 데이터 처리에 효과적이다. 또한, 과적합을 회피하여 모델 정확도를 향상시킨다. 그림 3은 Random Forest의 구조 및 동작 방식을 보여준다.

데이터 셋을 먼저 앙상블 할 모델의 개수로 나누고, 각 모델에 나눠진 데이터 셋을 학습시킨다. 그 후 모

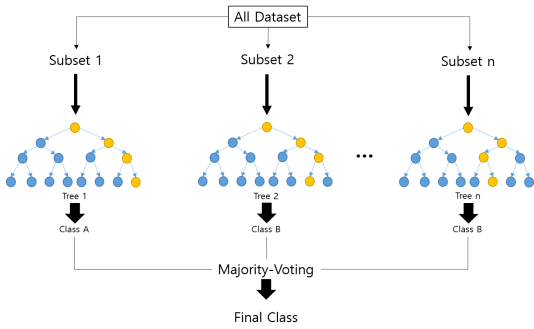


그림 3. Random Forest의 구조 및 동작 방식
Fig. 3. Structure and operation method of Random Forest

델들의 과반수 투표 알고리즘을 통해 최종 예측 결과를 내는 기법이다.

2.2.3 XGBoost(EXTreme Gradient Boosting)

XGBoost는 앙상블 기법 중 Boosting 계열에 속하는 알고리즘이다.^[6] Boosting 기법은 약한 모델을 순차적으로 생성하여 결과를 예측하는 기법으로, 이전 단계의 모델이 잘못 예측한 데이터를 다음 단계의 모델이 강하게 학습해서 순차적으로 생성된 모델들의 결과를 취합해서 최종 결과를 내는 기법이다. 그 중 Gradient Boosting 알고리즘은 경사 하강법(Gradient Descent)을 이용해 loss 함수 값을 최소화하여 파라미터를 최적화하는 방법인데, XGBoost는 Gradient Boosting 알고리즘을 분산 환경에서도 실행할 수 있도록 구현해놓은 라이브러리이다. 기본적인 Gradient Boosting에 비해 속도가 빠르고, 과적합 방지가 가능한 규제(Regularization)가 포함되어 있다. 또한, Early stopping을 제공해 과적합을 방지하면서 학습 시간을 줄일 수 있다. 그림 4는 Boosting 기법의 동작 방식을 보여준다.

먼저 데이터 셋을 나누고 모델에 넣어 결과를 보고 틀린 데이터의 가중치를 세게 하고 다음 모델이 학습

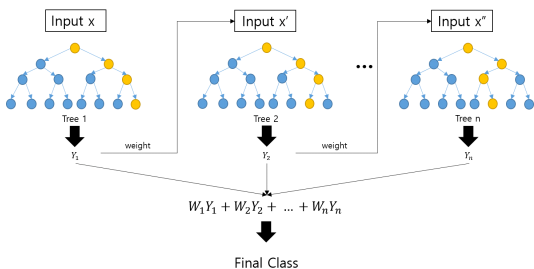


그림 4. Boosting 기법의 동작 방식
Fig. 4. Operation method of Boosting scheme

할 데이터 셋에 포함시켜 틀린 데이터에 대해 더 잘 학습할 수 있도록 한다. 이와 같은 과정을 여러 번 거쳐 각 모델들에 대한 예측 값들에 각 모델 별 가중치를 곱해 최종 결과를 내는 방식이다.

2.2.4 LightGBM(Light Gradient Boosting)

LightGBM은 Boosting 계열에 속하는 알고리즘으로 앙상블 동작 방식은 기존의 Boosting 앙상블 기법들과 동일하다.^[7] 이전에 XGBoost와 같은 GBDT(Gradient Boosting Decision Tree)의 구현체들이 있지만 효율성과 확장성 면에서 차원이 높거나 데이터 크기가 클 경우 여전히 만족스럽지 못한 결과를 보이는 경향이 있다. 또한, 데이터의 각 특징이 가능한 모든 분할 지점의 정보 획득량을 추정하기 위해 모든 데이터를 탐색해야 되기 때문에 비교적 시간이 오래 걸린다. 이러한 문제를 해결하기 위해 LightGBM에서는 GOSS(Gradient-Based One-Side Sampling)과 EFB(Exclusive Feature Bundling)을 제시하였다.

GOSS는 작은 기울기를 가진 데이터의 상당 부분을 제외하고, 나머지 부분만을 이용하여 정보 획득량을 추정하는 알고리즘이다. LightGBM은 GOSS를 통해 훨씬 더 작은 데이터 크기를 가지고 정확한 정보 획득량을 얻을 수 있다.

EFB는 특징의 개수를 감소시키기 위해 상호 배타적 특징을 묶는 기법이다. GBDT 모델에서는 feature scanning 알고리즘에서 feature histogram을 구축하는데, EFB를 통해 특징 묶음을 만들어 같은 방법으로 bundle histogram을 구축한다. 이 때, histogram 구축 복잡성은 데이터의 수를 N_D , 특징의 수를 N_F , 묶음의 수를 N_B 라고 할 때, $O(N_D \times N_F)$ 에서 $O(N_D \times N_B)$ 로 변경되어 기존의 GBDT 모델들보다 연산 양이 훨씬 줄어들게 된다.

또한 기존의 GBDT들은 트리를 구성할 때 균형을 우선하여 트리를 구성하였다. 그림 5에서 GBDT와 LightGBM의 트리 구성 과정을 통해 차이점을 보인다. 기존의 GBDT에서 사용하는 균형 우선 트리는 트리의 깊이가 줄어들거나 균형을 잡기 위한 연산이 추가된다는 단점이 있다. 하지만 LightGBM은 리프 노드를 지속적으로 분할하면서 트리를 구성해간다. 이와 같은 방법은 비대칭적이고 깊은 트리가 생성되지만 같은 리프 노드를 생성할 때 균형 우선 트리보다 손실을 줄일 수 있다는 장점이 있다.

LightGBM은 비대칭적인 트리를 생성하면서도 학습의 정확도에 영향을 끼치지 않는 새로운 방식을 찾아냄으로써 균형을 맞추는데 필요한 시간을 줄였다.

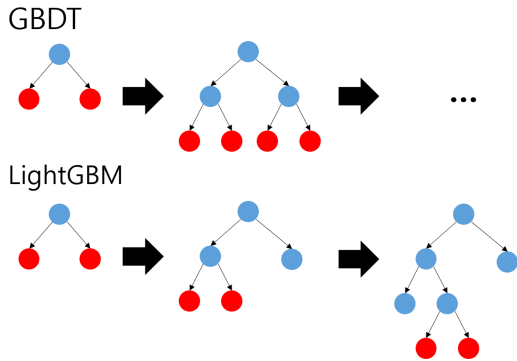


그림 5. GBDT와 LightGBM의 트리 구성 방식
Fig. 5. Tree organization method of GBDT and LightGBM

따라서 LightGBM은 대용량의 데이터 처리가 가능하며, 대용량의 데이터를 처리하면서도 기존의 GBDT 기법들에 비해 속도가 매우 빠르다. 또한, 다른 모델들 보다 적은 메모리를 사용한다는 장점을 가지고 있다.

2.2.5 기계 학습의 필요성

본 논문에서는 기계 학습을 이용하여 사용자의 위치, 요청 지연, 요청 패킷 크기 등의 다양한 정보를 학습하여 사용자의 요구 처리량 및 PRB 수에 따라 모델이 기지국을 선택하는 기법을 제안한다.

고정적인 알고리즘의 경우 실제 필드에서 사용자가 어떤 환경에 있는지 고려하지 않고 기지국의 입장에서만 판단하여 오프로딩 한다. 중간에 장애물이 없거나 주변 간섭이 없다면 좋은 결과를 나타낼 수도 있지만 사용자와 기지국 간 장애물이 있거나 간섭이 있다면 좋은 결과를 얻기 힘들다. 따라서 본 논문에서는 기지국이 실제로 사용자와 통신한 결과를 기반으로 모델을 학습시켜 매크로 셀 기지국이 주변 환경을 알게 하여 해당 통신 환경 내 최대한의 결과를 얻을 수 있도록 한다.

본 논문에서는 빠른 무선 환경 내에서 제한된 리소스를 활용해 실시간으로 데이터 처리가 필요하기 때문에 빠르면서도 자원을 적게 사용하는 LightGBM이 적합하다고 판단하였으며, 3장에서 실험을 통해 LightGBM이 적합한지 확인할 수 있었다.

2.3 데이터 오프로딩 여부 결정

본 논문에서는 매크로 셀 기지국에서 기계 학습을 이용하여 데이터 오프로딩 여부 결정을 위해 사용자의 매크로 셀 기지국과의 거리, 지연 민감도, 요청 패킷 크기 정보 등을 수집한다. 또한, 수집한 데이터에

따라 사용자에게 필요한 처리량을 계산하는 방법을 정의한다. 이 처리량에 따라 스몰 셀 기지국과 매크로 셀 기지국이 필요한 무선 PRB(Physical Resource Block)의 수를 계산하고, 사용자가 필요한 처리량을 높게 만족하면서 적은 수의 PRB를 사용할 수 있게끔 데이터 오프로딩을 결정하는 것이 본 논문이 제안하는 기법이다.

5G에서 사용자 $i, i \in N(N = 1, 2, \dots, n)$ 의 비트 크기와 지연, 그리고 처리량의 관계는 다음과 같다.

$$d_i = \frac{B_i}{T_i} \tag{3}$$

여기서, d_i 는 사용자의 요구 지연을 의미하고, B_i 는 사용자가 요구하는 패킷의 크기를 의미한다. 그리고 T_i 는 기지국에서 지원할 수 있는 사용자 i 의 처리량으로 다음과 같이 계산한다.^[8]

$$T_i = 10^{-6} \times Q_i \times R_{\max} \times \frac{b_i \times 12}{\tau_i} (1 - OH) \tag{4}$$

수식 (4)에서 b_i 는 대역폭이며, R_{\max} 는 LDPC 코드의 최대 수이고 OH 는 오버헤드 수, τ_i 는 OFDM 심볼 시간이다. Q_i 는 사용자 i 의 변조 방식에 따른 송신 비트 수로 비트 수가 높아질수록 송신 거리가 짧아진다. 그렇기 때문에 거리가 먼 사용자들은 높은 처리량을 받을 수 없다. 특히, Q_i 가 높아질수록 높은 경로 손실을 갖는 무선 채널에서는 오류(Error)가 많이 발생하여 사용자가 원하는 정보를 얻을 수 없다. 따라서 기지국에서는 각 사용자의 예상되는 경로 손실에 따라 Q_i 를 조절하여 보내는데, Q_i 가 낮아질수록 처리량이 낮아진다.

각 기지국이 가지고 있는 대역폭은 일정 범위로 정해져 있으며 이를 OFDM을 이용하여 나눈다. 나누어진 최소의 무선 자원 단위가 PRB이다. 즉, PRB를 사용자 i 에게 할당하는 수에 따라서 b_i 가 달라진다. 각 기지국에서 경로 손실이 높은 사용자에게 원하는 처리량으로 보내기 위해서는 Q_i 를 낮추고 많은 수의 PRB를 소모해야 한다. 특히, 스몰 셀의 경우 높은 주파수 대역을 가지고 있으므로 회절성이 약해 PRB를 많이 소모하여도 제대로 통신이 되지 않을 수 있다. 이런 상황일 경우 매크로 셀에서 데이터 오프로딩을 받아 처리할 수 있도록 해야 한다.

모든 네트워크의 경로 손실과 장애물의 유무 등을 알 수 있으면 가장 최적인 데이터 오프로딩을 할 수

표 1. 기계 학습을 이용한 기지국 간 데이터 오프로딩 알고리즘
 Table 1. Data Offloading Algorithm between Base stations with Machine Learning

Algorithm 1 : Data offloading between Base stations with Machine Learning	
1	User Number : N
	Input Data Set : $d_{3D}(SBS-X, SBS-Y, SBS-Z),$
2	$d_{3D}(MBS-X, MBS-Y, MBS-Z), d_{3D}(User-X,$ User-Y, User-Z), B_i, d_i, T_i
3	Data, Label = Labeling()
4	begin(Data, Label, New User Data Set) :
5	Model = CreatingModel()
6	Train(Model, Data, Label)
7	while Not the end of User do
8	Result = Predict(New User Data Set)
9	Data Offloading(Result)
10	end while
11	end

있다. 이는 최대 처리량이 가장 높으면서도 기지국이 PRB를 적게 소모하는 경우이다. 하지만 기지국이 모든 상황을 알 수 없으므로 본 논문에서 제안하는 기법은 이전에 처리 했던 사용자의 요구 지연, 패킷 크기, 각 기지국과의 거리 등을 기계 학습 기법으로 학습하여 스몰 셀 기지국과 매크로 셀 기지국 중 사용해야 하는 대역폭 즉, PRB를 적게 소모하면서도 높은 처리량을 받을 수 있는 기지국을 선택하도록 한다.

기지국 간에 데이터 오프로딩 알고리즘은 표 1과 같다. 먼저 이전 통신 결과에서 SBS와 MBS, 사용자의 위치, 원하는 요청의 지연, 데이터 크기 등을 데이터 셋으로 가져온다. 이 데이터를 3.1에서 서술할 방법에 따라 레이블링 한 후, 기계 학습 모델을 구축하여 학습한다.

마지막으로, 학습한 모델을 가지고 현재 새로운 사용자 요청을 어느 기지국이 처리할지 결정하는 알고리즘이다.

III. 실험

본 장에서는 실제 데이터에 대한 실험과 그 결과에 대한 분석을 함으로써 제안 기법에 대한 성능을 보여준다.

3.1 데이터 생성 및 레이블링

데이터는 시뮬레이터에서 네트워크 상황을 구축하여 시뮬레이터의 통신 결과와 네트워크 상태를 수집한 것으로 50000개의 MBS, SBS, 유저의 통신 결과이다. 본 논문의 기법은 오프로딩 최적화를 통해 총 처리량을 높이는 것이기 때문에 기존의 통신 기지국과는 별개로 어떤 기지국과 통신하였을 때 가장 높은 처리량을 갖는지에 대한 정보가 필요하다. 따라서 데이터를 보낸 유저의 설정을 기반으로 처리량을 계산하여 어떤 기지국과 통신하는 것이 가장 높은 처리량을 보이는지에 대한 레이블을 추가했다. 즉, 레이블의 생성 기준은 사용자의 위치 주변의 통신 결과 중에 가장 높은 처리량을 가질 수 있는 기지국을 선택하도록 한 것이다. 지도 학습의 분류 문제에서 레이블은 데이터에 대해 모델이 선택해야 할 정답이라고 할 수 있는데, 본 논문에서는 해당 위치에서의 다양한 요청에 대해 어느 기지국에서 처리하는 것이 가장 좋은지 선택하는 것에 대한 답이라고 할 수 있다. 데이터에 레이블을 매칭한 후 학습 데이터와 테스트 데이터로 나눴다. 학습 데이터와 테스트 데이터의 비율은 8 : 2로 무작위로 40000개와 10000개로 나누었다.

3.2 모델 평가 방법

기계 학습 모델은 다양한 방법으로 평가할 수 있는데 본 논문에서는 데이터 오프로딩 결정 여부를 분류 문제로 정의하여 풀어냈기 때문에 분류 모델에서 주로 사용되는 정확도, F1-score, ROC curve 및 AUC(Area Under the Curve)로 평가하였다. 본 절에서는 위의 3가지 평가 방법에 대해 서술한다.^{[9][10]}

3.2.1 정확도

정확도는 데이터에 따라 만든 레이블에 맞춰 모델의 예측이 레이블과 일치하는지 확인하여 맞은 개수를 세어 전체 데이터 개수에 대한 비율을 나타낸 것이다. 본 논문에서 정확도는 데이터 오프로딩 결정 여부에 대해 모델이 예측한 값과 레이블을 비교하여 일치한 데이터의 비율을 산출한 값이다. 식은 다음과 같다.

$$Accuracy = \frac{\#corrects}{\#data} \tag{5}$$

3.2.2 F1-score

F1-score는 Recall과 Precision의 조화 평균으로 데이터 레이블이 불균형을 이룰 때, 모델의 성능을 정확하게 평가할 수 있는 기준으로 정확도와 같이 성능을 하나의 숫자로 표현할 수 있다. 본 논문에서 Recall은

레이블링 과정에서 오프로딩 해야 한다고 판단한 데이터들에 대해 모델도 오프로딩 해야 한다고 판단한 데이터의 비율이며, Precision은 모델이 오프로딩 해야 한다고 판단한 데이터들에 대해 레이블도 그렇다고 판단한 데이터의 비율이다. 수식은 다음과 같다.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

3.2.3 ROC curve 및 AUC

ROC curve는 0-1 사이의 임계값에 따른 TPR(True Positive Rate)와 FPR(False Positive Rate)의 비율이다. 모델의 예측과 레이블을 비교해 보았을 때 TPR은 전체 데이터 수에 대해 예측과 레이블 모두 오프로딩 해야 한다고 분류한 데이터의 비율이고, FPR은 전체 데이터 수에 대해 예측은 오프로딩을 해야 한다고 분류했으나 레이블은 그렇지 않은 데이터의 비율이다. ROC는 그래프가 왼쪽 상단에 가까이 갈수록 좋은 결과이다.

AUC는 ROC 그래프의 면적을 측정한 결과로 ROC curve가 왼쪽 상단에 가까울수록 면적이 커지기 때문에 들은 비례한다고 볼 수 있다.

3.3 기계 학습 모델

학습 모델의 경우 Decision Tree, Random Forest, XGBoost, LightGBM과 앞의 모든 모델을 Bagging Ensemble한 모델의 5가지를 사용하여 비교해보았다. 모델들은 단일 트리 모델, Bagging Ensemble 모델, Boosting Ensemble 모델을 대표하는 모델들이다.^[11] 각각의 기법에서 대표적인 모델들에 데이터를 학습시켜 결과를 분석해본 후 데이터에 대해 최적으로 학습된 모델을 사용한다.

각 모델은 Scikit-learn 패키지의 Randomized Search CV를 통해 하이퍼 파라미터 튜닝을 진행했는데, 데이터의 특징 개수가 많지 않기 때문에 각 모델에서 성능에 영향을 많이 미치는 중요한 속성들만 튜닝하였다. 각 모델 별 튜닝 파라미터는 표 2와 같다.

본 논문에서는 10 folds 교차 검증을 통해 모델을 검증했다. 그림 6은 모델 별 10 folds 교차 검증 결과이다. 각 모델의 결과를 분석해보면 대체적으로 LightGBM이 잘 나오는 것을 볼 수 있다. LightGBM과 XGBoost가 대체적으로 상위에 위치해있는 것을 볼 수 있고, 그 뒤로 전체 모델을 합산한 Ensemble, Random Forest, Decision Tree 순으로 성능을 보인다. Decision Tree는 단일 모델로써 비교적 다른

표 2. 학습 모델과 파라미터
Table 2. Learning models and parameters

Model	Parameter
Decision Tree (CART)	Max_depths
	Max_features
	Min_samples_leaf
	Criterion
Random Forest	N_estimators
	Max_features
	Max_depth
	Criterion
	Min_samples_leaf
	bootstrap
XGBoost	Num_iterations
	Learning_rate
	Early_stopping_round
	Max_depth
	Boosting
LightGBM	Nrounds
	Eta
	Max_depth
	Loss
	Booster

Ensemble 기법에 비해 떨어지는 성능을 보인다. Random Forest 같은 경우, Decision Tree를 Bagging ensemble 한 기법으로 단일 Decision Tree에 비해 준수한 성능을 보이고 있어 Bagging Ensemble 기법이 단일 모델보다 확실히 좋은 정확도를 낸다는 것을 알 수 있다.

Boosting Ensemble 기법들로는 XGBoost와 LightGBM을 사용하였는데, XGBoost에 비해 LightGBM이 좋은 성능을 나타냈다. LightGBM이 XGBoost에 비해 좋은 성능을 보이는 것은 XGBoost에 비해 Generalization이 잘 되어 그런 것으로 보인다. LightGBM에서는 학습된 가중치에 따라 랜덤 샘플링하여 다음 학습 때의 가중치를 조절해 실패한 학습에 대해 잘 학습할 수 있도록 한다. 여기서 랜덤 샘플링 할 때의 과정이 Bagging과도 비슷해 Generalization에도 영향을 미치는 것으로 보인다. 실제로 트레이닝 셋의 정확도는 두 모델 모두 100%로 동일하다. 또한, LightGBM 모델은 XGBoost에 비해 모든 folds에 대해 비교적 균등한 정확도를 보인다. 이러한 결과를 통해 LightGBM 모델이 XGBoost 모델

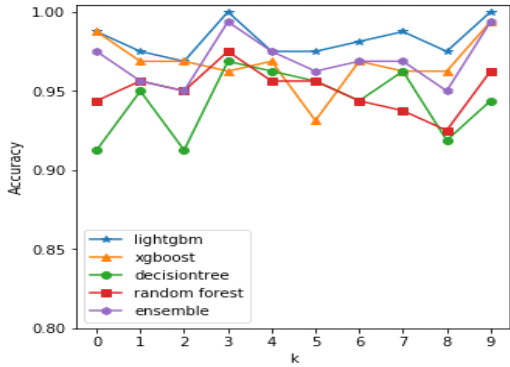


그림 6. 정확도를 통한 각 모델의 k-folds 교차 검증 평가
Fig. 6. k-folds cross validation of each model through accuracy

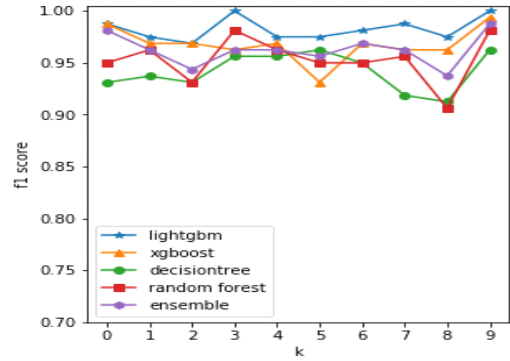


그림 8. F1 score를 통한 k-folds 교차 검증 평가
Fig. 8. k-folds cross validation through F1-score

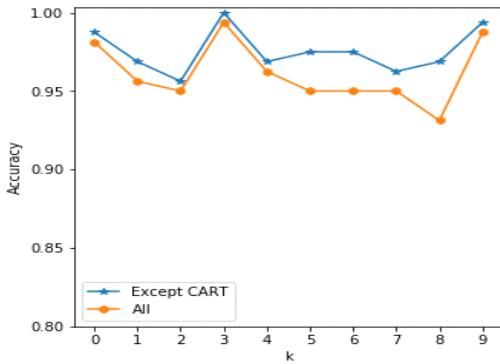


그림 7. 합성 모델의 종류에 따른 앙상블 모델의 정확도
Fig. 7. Accuracy of ensemble model by type of composed models

보다 Generalization이 잘 되었다고 볼 수 있다.

마지막으로 전체 모델을 모두 Bagging한 Ensemble 모델이 있다. 이 모델의 경우 위에서 사용한 LightGBM, XGBoost, Random Forest, Decision Tree를 모두 Bagging하여 Soft Ensemble 한 모델이다. 하지만 이 Ensemble 모델은 생각보다 좋은 성능을 내지 못하는데, 비교적 좋지 않은 성능을 가진 Decision Tree가 합쳐지면서 오히려 성능을 낮추는 것으로 보인다. 그림 7을 보면 Decision Tree를 제외한 Ensemble 모델이 모든 folds에 대해 높은 정확도를 보인다.

그림 8은 각 모델들의 F1-score 평가 기준을 적용한 그래프이다. 그림 6과 그림 8이 비슷한 그래프를 그리는 것으로 보아 데이터가 편향 없이 균등하게 분포되어 있음을 알 수 있다. 그림 9, 그림 10, 그림 11은 하나의 테스트 셋에 대한 결과로 정확도, F1-score, ROC curve 및 AUC로 나타난 모델 평가 그래프이다.

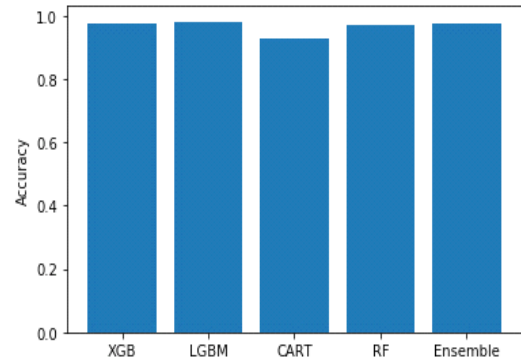


그림 9. 모델 별 테스트 셋에 대한 오프로딩 정확도
Fig. 9. Offloading accuracy of each model by test dataset

각각의 그래프에서 비교해보았을 때, LightGBM이 가장 상위에 위치한 것을 볼 수 있고 ROC curve 및 AUC도 0.99 이상의 매우 높은 점수를 보인다. 이와 같은 실험 방법들을 통해 Training set 및 Validation set을 학습 및 분석해본 결과 LightGBM을 선택한 것이 최적의 선택이었음을 알 수 있다.

3.4 네트워크 성능 평가

시뮬레이션 세팅은 표 3과 같이 설정하고 사용자의 위치는 무작위로 설정한다. N 은 무선 환경 내 존재하는 사용자 수이며 M 은 SBS의 수이다. Q_i 는 사용자 i 의 변조 방식에 따른 송신 비트 수이고, R_{\max} LDPC 코드의 최대 수이다. τ_s OFDM 심볼 시간이다. SBS Distance는 사용자와 가장 가까운 SBS와의 거리, MBS Distance는 사용자와 MBS와의 거리를 뜻한다.

각 그림에서 Proposed scheme은 본 논문에서 제시한 LightGBM을 활용해 무선 환경 내 통신 결과를 학습하여 매크로 셀로 오프로딩 하는 방법을 사용하였

표 3. 시뮬레이션 세팅
Table 3. Simulation setting

Parameter	Setting
N	50
M	3
Q_i	[2,8]
R_{max}	0.9257
τ_i	35μ
SBS Distance	[0,500]
MBS Distance	[0,1000]
SBS Centre Frequency	28GHz
MBS Centre Frequency	5GHz

을 때의 결과이다.

그림 12는 각 기법에 대한 전체 처리량을 나타낸다. Optimal은 무선 환경에 대한 모든 정보를 알고 있을 때 데이터 오프로딩 기준에 따른 시뮬레이션 결과이다. 기존의 거리에 따른 데이터 오프로딩 결정은 스몰 셀 기지국에서 많이 처리하여 스몰 셀 기지국에서는 처리량이 높으나 매크로 셀 기지국에서는 처리가 거의 이루어지지 않았음을 알 수 있다. 그에 비해 본 논문에서 제안한 기법은 매크로 셀 기지국으로 많은 요청들이 오프로딩 되어 스몰 셀 기지국이 과부하 되는 것을 방지할 수 있으며, 매크로 셀에서 바로 처리할 수 있는 사용자 요청을 오프로딩 하여 처리함으로써 같은 시간에 좀 더 높은 처리량을 보였다.

그림 13은 각 기법에 대한 전체 PRB 사용수이다. 기존의 기법에서는 스몰 셀 기지국이 대부분의 요청을 처리하여 스몰 셀 기지국에서 PRB를 많이 소모하며, 매크로 셀 기지국에서는 적게 소모하는 것을 확인할 수 있다. 그에 비해 제안한 기법에서는 스몰 셀 기

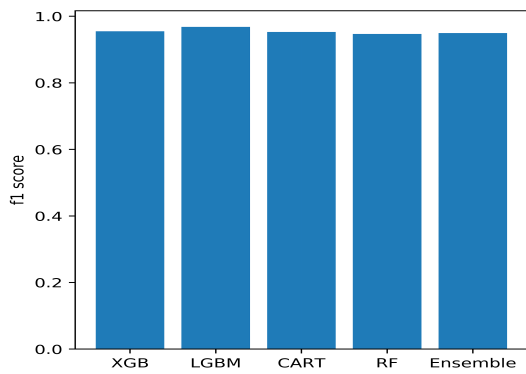


그림 10. 모델 별 테스트 셋에 대한 오프로딩 결정 F1 score
Fig. 10. F1 score of each model by test dataset

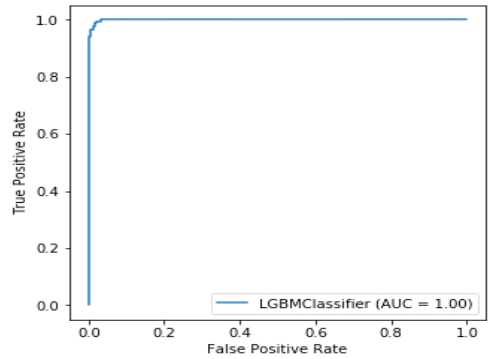


그림 11. ROC curve와 AUC를 통한 모델 평가
Fig. 11. Model evaluation through ROC curve and AUC

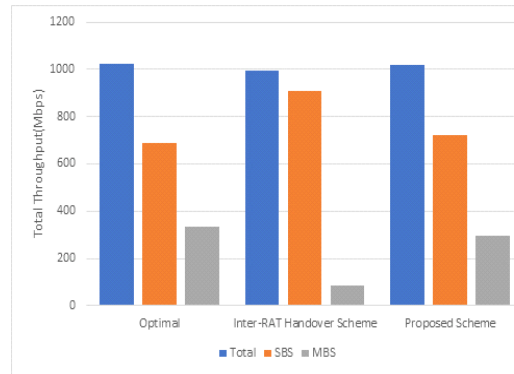


그림 12. 각 기법의 처리량
Fig. 12. Throughput of each scheme

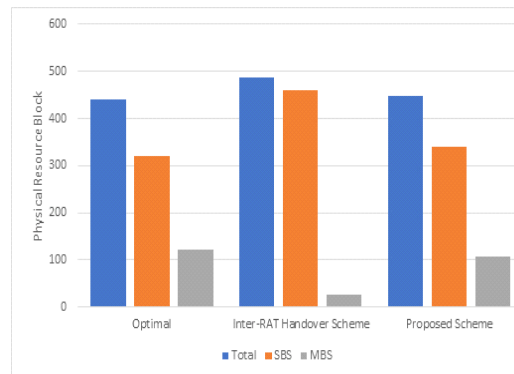


그림 13. 각 기법의 PRB 사용 수
Fig. 13. The number of PRBs used by each scheme

지국의 PRB 소모를 줄인 것을 확인할 수 있다. 이에 따라 스몰 셀 기지국을 위해 매크로 셀 기지국에서 스몰 셀에 오프로딩을 요청하여 사용자 요청을 로드 밸런싱 하였다. 또한, 기존의 기법보다 본 논문에서 제

안한 기법에서 전체 PRB 소모 수가 적은 것을 확인할 수 있다. 위와 같은 실험 결과를 통해 제안한 기법이 무선 자원의 소모를 줄이면서 전체 시스템 성능을 높였음을 알 수 있다.

IV. 결 론

스몰 셀 기술은 밀리미터 대역을 지원하기에 적절한 기술이지만 좁은 범위의 통신으로 인해 많은 요청을 처리하기에는 과부하 문제가 생긴다는 단점을 가지고 있다.

본 논문에서는 네트워크를 스몰 셀 기지국과 매크로 셀 기지국으로 나눈 후 스몰 셀 기지국에서 생기는 작업 과부하 문제를 해결하기 위해 매크로 셀 기지국에서 기계 학습을 이용하여 사용자들 데이터 오프로딩 하는 알고리즘을 제시한다. 데이터 오프로딩 기준으로는 사용자들과 각 셀 기지국의 거리, 지연 민감도, 요청 패킷 크기 등의 실제 통신 결과를 머신러닝 모델을 이용해 학습하여 매크로 셀 기지국과의 통신 적합성을 기준으로 가장 적합한 사용자들을 매크로 셀 기지국으로 데이터 오프로딩 하여 전체 성능을 높일 수 있음을 보였다.

또한, LightGBM 모델을 활용하여 다른 머신러닝 알고리즘에 비해 적은 학습 시간과 지연 시간을 가진다. 기존 알고리즘에서 사용하는 XGBoost의 경우 연산 복잡도가 $O(\#data \times \#feature)$ 이지만 LightGBM의 경우 $O(\#data \times \#bundle)$ 로 특징을 묶어 훨씬 더 적은 연산 복잡도를 가진다. 또한, 그러면서도 더 정확한 분류 결과를 얻을 수 있었으며 기존의 거리만을 활용하는 알고리즘에 비해 높은 성능을 얻을 수 있었다.

향후에는 더 많은 무선 환경에 대한 정보와 많은 사용자에게 대한 학습을 위하여 심층 학습(Deep Learning)을 활용하는 연구 등이 활발해질 것으로 예측된다. 또한, 매크로 셀 기지국과 스몰 셀 기지국뿐만 아니라 스몰 셀 기지국과 스몰 셀 기지국 간, 기지국과 Wi-Fi 등의 다양한 통신 주체 간의 데이터 오프로딩 결정으로 더욱 성능을 높일 것으로 생각된다. 또한, 오프라인 학습 방법으로 학습한 머신러닝 모델은 학습 시간이 오래 걸릴 수 있어 환경의 변화에 즉각 대응하기 힘들기 때문에 강화학습 모델이나 신경망의 온라인 학습 방법 등 급작스런 환경의 변화를 대응할 수 있는 기법에 대한 연구가 필요할 것으로 보인다.

References

- [1] H. Park, M. Kang, and S.-W. Jeon, "5G/6G channel characteristics and channel model analysis," in *Proc. Symp. KICS*, pp. 529-530, 2020.
- [2] F. B. Mismar and B. L. Evans, "Partially blind handovers for mmWave new radio aided by sub-6 GHz LTE signaling," *2018 IEEE ICC Wkshps.*, pp. 1-5, Kansas City, MO, 2018.
- [3] 3GPP, *Study on channel model for frequencies from 0.5 to 100 GHz*, TS 38.901 V16.1.0, System Architecture for the 5G System (Release 16, 2019), Retrieved Jul. 30, 2020, from <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3173>
- [4] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, pp. 81-106, 1986.
- [5] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, pp. 5-32, 2001.
- [6] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD '16*, San Francisco, CA, USA, Aug. 2016.
- [7] G. Ke, et al., "Lightgbm: A highly efficient gradient boosting decision tree," *NIPS 2017*, pp. 3146-3154, Dec. 2017.
- [8] 3GPP, *NR; User Equipment (UE) radio access capabilities*, TS 38.306 V16.2.0, System Architecture for the 5G System (Release 16, 2020), Retrieved Jul. 31, 2020, from <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3193>
- [9] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, pp. 29-36, 1982.
- [10] Y. Sasaki, "The truth of the F-measure," *Teach. Tutor. Mater.*, Oct. 2007.
- [11] P. Bühlmann, *Bagging, Boosting and Ensemble Methods*, Handbook of Computational Statistics, 2012.

안 상 현 (Sanghyeon An)



2019년 2월 : 한림대학교 컴퓨터
공학과 졸업
2019년 3월~현재 : 서강대학교 컴
퓨터공학과 석사 과정
<관심분야> 무선 통신, 머신 러
닝

전 우 선 (Woosun Chon)



2019년 2월 : 한신대학교 정보통
신학과 졸업
2019년 3월~현재 : 서강대학교
컴퓨터공학과 석사 과정
<관심분야> 무선 통신, 머신 러
닝

김 종 원 (Jongwon Kim)



2019년 2월 : 동양미래대학교 정
보통신과 졸업
2019년 3월~현재 : 서강대학교
컴퓨터공학과 석사 과정
<관심분야> 무선 통신, 머신 러
닝

소 정 민 (Jungmin So)



2006년 8월 : University of Illinois
at Urbana-Champaign 박사
2006년 9월~2010년 2월 : 삼성
전자 책임연구원
2011년 3월~2018년 8월: 한림대
학교 컴퓨터공학과 부교수
2018년 9월~현재 : 서강대학교
컴퓨터공학과 교수

<관심분야> 무선네트워크, 머신러닝

[ORCID:0000-0001-5486-9068]