

데이터 증강 기반 효율적인 무선 신호 분류 연구

¹*임상순

An Efficient Wireless Signal Classification Based on Data Augmentation

¹*Sangsoon Lim

요약

사물인터넷 환경에서는 다양한 무선 통신 기술을 사용하는 기기들이 점점 증가하고 있다. 특히, 다양한 무선 신호 변조 유형을 정확하게 식별하기 위해 효율적인 특성 추출 기법을 설계하고 무선 신호의 종류를 분류하는 것이 필수적이다. 하지만, 실제 환경에서 레이블이 지정된 무선 신호 데이터를 수집하는 것은 쉬운 문제가 아니다. 최근 무선 신호 분류를 위해 딥러닝 기반의 다양한 학습 기법들이 제안되어졌다. 딥러닝의 경우 훈련 데이터셋이 적을 경우 과대적합에 빠질 가능성이 높으며, 이는 딥러닝 모델을 활용한 무선 신호 분류 기법의 성능 저하를 유발한다. 본 연구에서는 다양한 무선 신호들이 존재할 때 분류 성능을 높이기 위해 생성적 적대 신경망 기반 데이터 증대 기법을 제안한다. 분류해야 하는 무선 신호의 종류가 다양할 때 특정 무선 신호를 나타내는 데이터의 양이 적거나 균형이 맞지 않는 경우 제안한 기법을 활용하여 필요한 무선 신호와 관련된 데이터의 양을 증가시킨다. 제안한 데이터 증강 알고리즘의 유효성을 검증하기 위해 무선 신호의 데이터양을 증가시키고 균형을 맞춘 결과를 바탕으로 CNN 및 LSTM 기반 무선 신호 분류기를 구현하여 실험해본 결과 데이터 균형을 맞추지 않았을 때보다 분류 정확도가 높아지는 것을 확인하였다.

Abstract

Recently, diverse devices using different wireless technologies are gradually increasing in the IoT environment. In particular, it is essential to design an efficient feature extraction approach and detect the exact types of radio signals in order to accurately identify various radio signal modulation techniques. However, it is difficult to gather labeled wireless signal in a real environment due to the complexity of the process. In addition, various learning techniques based on deep learning have been proposed for wireless signal classification. In the case of deep learning, if the training dataset is not enough, it frequently meets the overfitting problem, which causes performance degradation of wireless signal classification techniques using deep learning models. In this paper, we propose a generative adversarial network(GAN) based on data augmentation techniques to improve classification performance when various wireless signals exist. When there are various types of wireless signals to be classified, if the amount of data representing a specific radio signal is small or unbalanced, the proposed solution is used to increase the amount of data related to the required wireless signal. In order to verify the validity of the proposed data augmentation algorithm, we generated the additional data for the specific wireless signal and implemented a CNN and LSTM-based wireless signal classifier based on the result of balancing. The experimental results show that the classification accuracy of the proposed solution is higher than when the data is unbalanced.

Keywords: Wireless Signal, Signal Classification, Deep learning, GAN, IoT

¹* 성결대학교 컴퓨터공학과 조교수 (lssgood80@gmail.com)

I. 서론

사물인터넷 환경에서는 다양한 무선 통신 기술을 사용하는 기기들이 점점 증가하고 있다[1]. 또한, 정보 기술의 급속한 발전과 함께 무선 신호 유형 분류 및 변조 방식 식별을 포함한 신호 분류는 항공기 통신 및 기타 분야의 무선 자원 관리에서 점점 더 중요한 역할을 하고 있습니다[2]. 이러한 환경에서 무선 기기들 간의 상호 간섭을 최소화하기 위해 주변에 존재하는 무선 신호들의 특성을 파악하고 최대한 신호를 분류하여 정보를 제공할 수 있는 상호 협력 가능한 환경 구축에 대한 필요성이 증가하고 있다[3]. 하지만, 서로 다른 통신 기술들이 존재하는 환경에서는 다양한 간섭이 발생하고 이를 응용하는 서비스들의 특성에 따라 고정적인 규칙 기반의 무선 신호 탐지 및 분류 기술은 효율적이지 못하다는 단점이 있다.

기존의 무선 신호 특성을 분류하기 위한 연구에서는 특정 무선 기술의 패킷 길이, 패킷 전송 간격, 패킷 전송 유지 시간, 패킷의 수신된 신호 세기 특성 등을 고려하여 Wi-Fi, Bluetooth, ZigBee, 전자레인지 등 2.4GHz 대역에 동작하는 다양한 무선 기술들을 판단할 수 있는 기법들을 제안하였다[4][5]. 이러한 규칙 기반의 무선 신호 분류 기술은 주변에서 발생하는 환경 잡음에 따라 그 특성이 쉽게 변할 수 있으며 신호를 발생시키는 기기와 신호를 수집하는 기기 사이의 위치 변화에 따른 간섭 형태의 변화에도 민감하게 영향을 받는 문제로 인해 현실적인 분류 정확도를 높이는 데 한계가 존재한다.

최근에는 무선 신호들의 특성을 추출하여 자동으로 무선 신호의 종류를 분류하는 딥러닝 기술들이 제안되어졌다[2][6][7]. 기존의 지도 학습을 활용하는 기법들의 경우 많은 양의 레이블이 지정된 데이터가 필요하다. 특정한 환경에서 발생하는 무선 신호들을 체계적으로 분류하기 위해서는 해당 환경의 특성이 적용된 실제 측정 데이터가 충분히 수집되어야 하지만 해당 환경에 설치되어 있는 무선 신호 수집기가 다양하지 않기 때문에 활용 가능한 다양한 고품질의 무선 신호 데이터가 매우 부족한 상황이다. 따라서, 특정 무선 신호에 대한 샘플의 수가 적은 불균형한 데이터셋을 기반으로 딥러닝을 활용한 무선 신호 분류기를 구성했을 때 레이블이 지정된 데이터가 부족하여 딥러닝 모델의 훈련 과정이 쉽게 과적합되고 분류 성능이 낮아지는 문제가 발생한다.

생성적 적대 신경망(Generative Adversarial Network, GAN)은 2014년 Goodfellow에 의해 제안되었다[8]. GAN의 기본적인 동작 방식은 강력한 비선형 피팅 기능을 사용하여 임의의 사전 노이즈 분포에서 실제 데이터 분포로의 비선형 매핑을 학습하여 생성기가 실제 샘플을 생성할 수 있도록 하는 것이다. GAN은 이미지 생성, 비디오 처리, 음성 생성, 텍스트 처리 및 기타 여러 딥러닝 활용 분야에서 널리 사용되고 있다.

본 논문에서는 GAN을 활용하여 데이터의 불균형이 심화되어 있는 환경에서 데이터가 균형을 이룰 수 있도록 적은 데이터를 증대시키는 기법을 제안한다. 무선 신호의 데이터는 I, Q 값을 이미지화 하여 나타낼 수 있으므로 분류해야 하는 무선 신호들의 종류 중 부족한 신호 샘플에 대해 GAN을 통해 증대시킬 경우 데이터셋 전체적인 균형을 맞출 수 있는 특징이 있다. 각 무선 신호 종류별로 균형있게 구성된 데이터셋을 CNN(Convolutional Neural Network) [9] 과 LSTM[10][11] 레이어가 혼합된 무선 신호 분류 모델을 통해 학습시키고 분류를 진행한 결과 데이터의 불균형이 존재하는 경우에도 모델의 정확도를 향상시킬 수 있음을 확인하였다.

본 논문의 2 장에서는 기존에 연구되었던 무선 신호 분류와 관련된 기법들에 대해 소개한다. 3 장에서는 제안하고자 하는 데이터 불균형 문제 해소를 위한 데이터 증대 기법과 이를 적용한 딥러닝 신호 분류 기법을 설명한다. 4 장에서는 데이터 증대 결과와 이를 적용한 딥러닝 모델의 성능 검증 결과를 보여주며, 5 장에서 결론을 맺고 논문을 마무리한다.

II. 관련연구

2.1 규칙 기반 무선 신호 분류

규칙 기반 무선 신호 분류에서 가장 기본적인 접근 방식은 수신된 신호 세기의 패턴을 분석하는 방법이다. 다양한 무선 신호들에 대해 수신된 신호 세기를 수집하고 수신된 신호 세기의 특징들을 규칙 기반으로 분류하게 된다.

Rayanchu[5]는 2.4GHz 대역에서 동작하는 Wi-Fi 인터페이스를 활용하여 Wi-Fi 이외의 다른 무선 기술들의 신호를 발견하는 기술을 제안하였다. Wi-Fi 인터페이스는 상대적으로 다른 2.4GHz 대역의 무선 인터페이스보다 빠른 속도로 무선 신호를 수집하고 처리할 수 있는 능력이 있으며, 이를 바탕으로 주변 무선 신호를 수집하여 해당 신호들의 신호 세기를 정교하게 분석하였다. 해당 기술에서는 빠르게 Wi-Fi 이외의 무선 기술을 사용하는 기기들을 발견하고 위치를 특정하여 간섭을 주는 강도를 분석할 수 있는 기법을 제공한다. 무선 신호를 수집하고 신호 동기화 및 클러스터링을 수행한 후 신호의 특징들을 규칙 기반으로 분류하고 간섭의 정도와 해당 신호를 발생시키는 위치를 특정할 수 있도록 설계되었다.

또한, 일반적인 가능성 기반 자동 무선 변조 기술 분류(Automatic Modulation Classification, AMC)는 수집된 신호들에 대해 전체 변조 신호의 종류들에 대한 가능성 비율을 결정하고 가능성 높은 변조 후보를 얻기 위한 임계값을 고려한다. 이러한 무선 변조의 기본적인 특성 관련 접근 방식은 AMC를 규칙 기반의 패턴 감지 문제로 다루게 된다[12]. 즉, 수집된 신호에서 서로 다른 특징을 결정하고 이 특징의 분석된 값에 따라 분류가 실행되므로 가능성 기반 접근 방식은 높은 분류 정확도 결과를 제공한다.

하지만, 규칙 기반 무선 신호 분류와 같은 접근 방식에서는 결과 도출을 위해 진행되는 연산들의 복잡도가 증가하는 문제로 인해 빠른 분류가 진행되어야 하는 환경에서는 적합하지 못한 단점이 있다.

2.2 딥러닝 기반 무선 신호 분류

최근 딥러닝 기술을 다양한 분류 시스템에 적용되는 사례가 늘어나고 있다. 무선 신호의 경우에도 종류별로 명확히 구분되는 특징들이 있으며 이러한 특징들을 딥러닝 접근 방법을 통해 학습하고 도출된 모델을 바탕으로 자동으로 구분해 줄 수 있는 기법들이 제안되었다.

Zhou[13]는 자동적으로 변조 기법을 분류하기 위해 이미지 처리에서 주로 사용되던 CNN 기법을 적용하는 것을 제안했다. 대표적인 15 개의 무선 신호 종류에 대해 신호의 특성을 CNN 모델로 학습하여 새롭게 발생하는 신호에 대해 자동으로 분류를 진행하는 방식으로 규칙 기반 기법과는 달리 별도의 속성을 추출하지 않고 직접적으로 신호를 분류할 수 있는 장점이 있으며, 필요에 의해 수동으로 특징을 추출할 수도 있다. 전체 수집된 신호 중 특정 신호 대 잡음비(Signal to Noise Ratio, SNR) 범위에서 뚜렷하게 도출되는 특성들을 바탕으로 모델을 생성하며, 낮은 신호 대 잡음비에서도 신호를 구분할 수 있는 것을 알 수 있다.

Zhang[14]은 추가적인 백색 가우시안 잡음(Additional White Gaussian Noise, AWGN) 환경에서 신호 대 잡음비가 낮은 조건에서 특정 신호의 변조 인식 정확도가 낮은 문제를 해결하기 위해 차동 비선형 위상 피크 팩터와 반복적인 진폭의 차이를 추가로 고려할 수 있는 특성들을 기존 AMC 기법에 추가하는 방법을 제안하였다. 11 개의 변조 신호를 구분하기 위한 특징 선택 알고리즘 기반의 트리 형태의 다층 서포트 벡터 머신 분류기를 제안하였다.

Hong[15]은 수집한 무선 신호의 시계열 특성을 적절히 활용할 수 있도록 RNN(Recurrent Neural Network) 기반의 새로운 자동 변조 분류 기법을 제안하였다. 이 기술은 제한된 데이터 길이에 의해 원시 신호에 직접 의존하고 신호 기능 추출 없이 분류할 수 있도록 지원한다. 제안하는 방법을 CNN 기반 방법과 비교한 결과, 신호 대 잡음비가 -4dB 이상인 상황에서 좋은 성능을 나타내는 것을 알 수 있었다. 또한, 2 계층 GRU(Gated Recurrent Unit) 네트워크를 구성했을 때 효율성이 높은 부분을 검증하였다.

하지만, 딥러닝 모델 성능의 경우 수집한 데이터셋의 구성이 정확도에 큰 영향을 줄 수 있다. 특히, 데이터셋에서 일부 신호의 데이터양이 부족하여 데이터가 불균형이 존재할 때는 정확도가 낮아지는 문제가 발생하게 된다. 본 논문에서는 기존 연구 방식들 중 효율성이 높은 딥러닝 기반 무선 신호 분류 시스템의 데이터 불균형 상황에서의 단점을 극복하기 위해 데이터 증강 기법을 활용하는 방안을 제안한다. 분류해야 하는 신호의 개수가 많은 경우 각각의 신호에 대한 특성들을 충분히 학습할 수 있을 만큼의 데이터셋을 확보하는 것이 전체적인 성능에 큰 영향을 주게 된다. 따라서, 전체 신호들에 대한 데이터셋 중 상대적으로 적은 양의 데이터셋을 확보하고 있는 무선 신호의 데이터를 증강시켜 전체적인 데이터 균형을 조절하여 성능을 향상시킨다.

III. 데이터 증강 기반 무선 신호 분류

3.1 생성적 적대 네트워크(GAN)와 데이터 증강

GAN은 서로 대립하는 생성자(Generator)와 판별자(Discriminator)로 구성된다. 생성기 G는 실제 표본의 확률 분포를 추론하고 생성된 표본의 확률 분포가 실제 표본의 확률 분포에 근사하도록 만든다. 판별기 D는 샘플이 데이터셋의 실제 샘플인지 또는 생성기 G에서 생성된 가짜 샘플인지를 구별한다. 생성기 G는 임의의 노이즈를 입력으로 사용하여 생성된 샘플을 출력하는 반면 판별기 D의 입력은 두 소스(실제 샘플과 생성기에 의해 생성된 샘플)에서 나온다. 그리고 샘플의 두 부분의 진위를 구별하려고 시도한다. 생성자 G와 판별자 D는 최종적으로 수렴하여 내쉬 평형(Nash Equilibrium)에 도달할 때까지 동적 게임을 형성하도록 교대로 훈련된다. 여기에서의 내쉬 평형은 생성기의 데이터가 실제 데이터에 매우 가깝고 판별자가 이를 올바르게 구별할 수 없는 상태를 의미한다. GAN의 전체 프로세스와 관련된 손실 함수는 다음과 같이 표현할 수 있다.

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

여기서 p_{data} 는 실제 데이터 분포이고, p_z 는 잡음 분포를 나타내며, \sim 은 분포 준수를 나타내고, E 는 기대 연산자를 나타내고, $V(D, G)$ 는 GAN의 결합된 목적 함수를 나타내고, $D(x)$ 는 D 가 실제 데이터에 건전한 차별성을 부여할 확률이다. $G(z)$ 는 G 가 생성한 가짜 데이터를 의미하며, $D(G(z))$ 는 D 가 생성된 가짜 데이터를 실제 데이터로 판단할 확률을 나타낸다. 생성기 G 의 경우 훈련 목표는 매개변수를 조정하고 $\log(1 - D(G(z)))$ 를 최소화하는 것이다. 즉, 생성된 데이터 분포가 실제 데이터 분포에 가까울수록 생성기의 성능이 더 좋은 것을 의미한다. 판별기 D 의 경우 훈련 목표는 매개변수를 조정하고 $\log(D(x))$ 를 최대화하고 $\log(D(G(z)))$ 를 최소화하는 것이다. 즉, 판별 능력이 강할수록 판별기의 성능이 더 좋다는 것을 의미한다. 따라서 G 와 D 는 최종 수렴까지 최적화 과정에서 서로 게임을 진행한다. 여기서 D 는 실제 데이터와 생성된 데이터의 차이를 구분할 수 없으며 G 는 더 이상 생성된 샘플의 품질을 개선할 수 없는 경우 내쉬 균형에 도달했다고 판단한다.

[그림 1]은 GAN 기반 무선 신호 데이터 증강 구조를 나타낸다. 학습 과정에서 랜덤 노이즈 신호가 생성기에 공급되고 판별기는 실제 샘플과 생성된 샘플을 구별한다. 생성기와 식별기는 동시에 학습이 진행되고 매개변수를 공유하여 입력 신호와 유사한 신호 샘플을 생성한다. 본 연구에서는 GAN의 성능을 보다 현실적으로 판단하기 위해 무선 신호 분류 논문들에서 가장 많이 활용되고 있는 무선 신호 샘플인 RML2016.10a[16] 데이터셋을 기본 데이터로 활용하였다. 실제 데이터셋에서 주로 활용되는 무선 신호에 대해 데이터 증강을 수행한 결과 실제 신호와 매우 유사하지만 약간씩 다른 변조된 신호가 생성되는 것을 확인할 수 있었으며, 해당 신호들의 변조 기법에 부합하는 특징을 갖는 것을 확인하였다. 즉, GAN 기반 데이터 증강 기술을 활용하여 데이터셋의 다양성을 늘릴 수 있으며 이를 딥러닝 훈련에 직접적으로 사용가능한 것을 확인하였다.

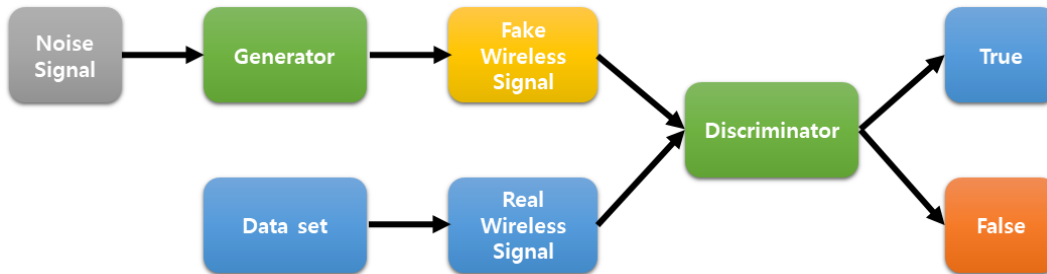


Figure 1. The Structure of Wireless Signal Data Augmentation based on GAN
그림 1. GAN 기반 무선 신호 데이터 증강 구조

3.2 딥러닝 기반 무선 신호 분류 기술

본 연구에서는 딥러닝 기반 무선 신호 분류 기술을 위해 RadioML2016.10a 데이터셋을 활용한다. RadioML2016.10a 데이터셋은 GNU Radio 기반으로 생성된 무선 신호 데이터셋으로 다양한 무선 신호 분류 연구에서 활용되어 왔다. 일반적으로 데이터셋에 있는 원시 신호는 다양한 신호 대 잡음비 환경에서 존재할 수 있는 $I = A\cos(\phi)$ 및 $Q = A\sin(\phi)$ 값 들로 구성되어진다. 여기서 A 및 ϕ 는 수집된 신호의 순간 진폭 및 위상을 나타낸다. RadioML2016.10a 데이터셋은 실제 다중 경로 페이딩 및 백색 잡음이 존재하는 환경을 구성하여 일반적으로 많이 사용되는 신호 변조 방법들 11 가지에 대해 신호를 데이터셋을 구축하였다. 11 가지 변조 신호의 종류는 8PSK, AM-DSB, AM-SSB, BPSK, CPFSK, GFSK, 4PAM, 16QAM, 64QAM, QPSK 과 WBFM 이다. 이 데이터셋에는 심볼당 4 개의 샘플 신호가 포함되어 있으며 가 128 개 단위의 샘플 데이터가 포함되어 있다. 또한, 파이썬 피클 파일 형태로 저장된 디서너리형 데이터로 키와 값 형태로 구성되어 있다. 각 키는 11 개의 변조 방식과 -20~18dB 의 SNR 튜플로 구성되며 값은 220 개의 키 값에 해당하는 (1000, 2, 128)의 넘파이 배열을 갖는다. 즉, 128 개 샘플의 I 및 Q 두 값이 있는 1000 개의 샘플 윈도우로 구성된다.

본 연구에서는 정확도 높은 무선 신호 분류기를 설계하기 위해 CNN 과 LSTM 을 혼합하여 모델을 설계하였다. CNN 은 입력 데이터가 컨볼루션 레이어와 풀링 레이어를 거치도록 구성하고 마지막에 완전 연결 레이어를 통과하는 구조를 가지고 있다. 컨볼루션 레이어는 입력 데이터의 형태를 유지하고 각 필터에 대한 특징 맵을 생성하는 역할을 하며, 풀링 레이어는 컨볼루션 레이어의 출력 데이터를 입력으로 받아 출력 데이터의 크기를 줄이거나 특정 데이터를 강조하는 역할을 한다. 이를 통해 전체 데이터 중 일부 왜곡된 데이터를 추출할 수 있는 장점이 있다. 또한, 이 레이어는 속성을 자동으로 추출할 수 있으며 컨볼루션 레이어는 필수적으로 필요하며 풀링 레이어는 훈련 형태에 따라 다양하게 구성 가능한 특징이 있다. 무선 신호 데이터의 경우에도 I, Q 값에 대해 CNN 을 통해 주요 특징을 도출하여 분류의 정확도를 높일 수 있다.

일반적인 컨볼루션 레이어는 $N \times M$ 형태의 2 차원 행렬 필터를 사용하여 이미지의 특징을 추출한다. 높이와 너비로 구성된 2 차원 데이터에서 지정된 간격으로 $N \times M$ 크기의 필터를 순회하면서 중첩된 데이터에 커널의 요소 값을 곱한 값을 모두 합산하는 방식이다. 이때, 이동 간격을 스트라이드라고 하며, 스트라이드를 1 로 지정하면 한 열씩 이동하며 컨볼루션 과정을 진행한다. 컨볼루션의 출력을 특징 맵이라고 하고 활성화 함수를 특징 맵에 적용한 것을 활성화 맵이라고 한다. 컨볼루션 과정을 거친 후 출력 데이터는 입력 데이터보다 작아지는데 패딩이라는 과정을 거쳐서 출력 데이터가 줄어드는 것을 방지할 수 있다. 패딩은 일반적으로 0 인 지정된 크기만큼 특정 값으로 입력 데이터의 가장자리를 채우는 과정을 의미한다. 풀링 계층에는 최대 풀링, 최소 풀링 및 평균 풀링이 포함될 수 있다. 컨볼루션 레이어와 마찬가지로 필터와 스트라이드 개념이 풀링 연산에도 적용될 수 있다. 최대 풀링은 필터와 데이터가 겹치는 영역에서 최대값을 추출하는 기법이고 평균 풀링은 겹치는 영역의 평균을 추출하는 기법이다. 컨볼루션과 풀링은 필터와 스트라이드 개념을 사용한다는 점에서 비슷해 보이지만 풀링 연산은 가중치가 없다는 점이 다르다.

또한, 기존 딥러닝 연구에서는 시간적인 상관 관계가 있는 시계열 데이터를 학습시키기 위해 RNN 을 활용한다. RNN 은 과거에서 미래로 정보를 전파할 수 있는 내부 메모리가 있는 기존 피드포워드 네트워크의 일반화된 형태이다. 이는 정보가 지속되도록 하는 네트워크에서 루프를 생성하고 일련의 입력 데이터를 처리하기 위해 메모리 상태 정보를 동시에 활용하는 구조이다. RNN 은 정보 저장소 역할을 하는 숨겨진 계층을 유지하면서 이전 출력을 활용할 수 있다. 이는 동일한 가중치와 편향이 모든 레이어에 제공되므로 매개변수를 기억하고 늘리는 문제가 최소화되지만 연결된 노드가 길어질수록 가중치가 역방향으로 전달되는 과정에서 기울기가 소실되는 문제가 있다.

LSTM 은 게이팅 메커니즘을 통해 장기 종속성을 학습하는 RNN 의 고급 버전이다. 또한, 기존 RNN 에서 훈련할 때 발생하는 기울기 소실 문제에 대한 해결할 수 있다. LSTM 은 게이트와 셀 상태를 가지고 있으며, 셀 상태는 네트워크의 메모리로 동작하며 관련 정보를 전파하는 경로 역할을 한다. 즉 망각, 입력, 출력 게이트는 정보가 전달되는 흐름을 제어하며 유지하거나 폐기할 지식을 결정하고 이 과정에서 기존에 발생하는 기울기 소실 문제를 완화시킬 수 있다. [그림 2]는 무선 신호를 높은 정확도로 분류하기 위해 설계한 CNN 과 LSTM 혼합 모델을 보여준다. 수집된 데

이더넷에 대해 Conv1D 레이어를 통과시킨 후 풀링과 LSTM 과정을 두 번 연속적으로 수행하여 무선 신호 데이터를 학습함으로써 90% 이상의 정확도를 보이는 분류 모델을 생성할 수 있다.

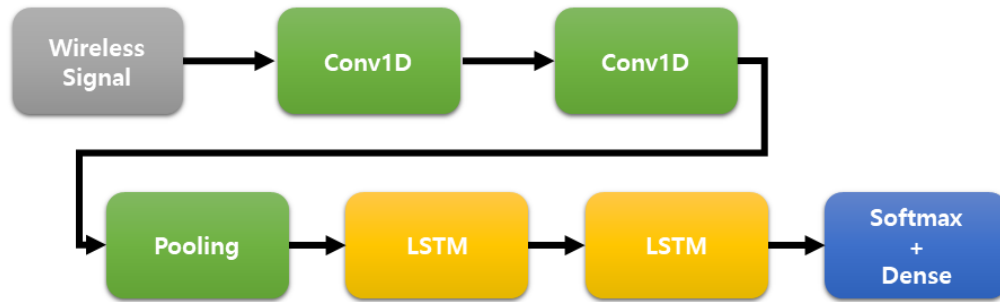


Figure 2. The Structure of Deep learning Model for Wireless Signal Classification
그림 2. 무선 신호 분류를 위한 CNN/LSTM 혼합 딥러닝 모델

이때, 수집한 데이터셋의 무선 기술 종류별 신호 데이터에 대해 각 기술별로 동일한 비율의 데이터가 존재할 때는 성능에 문제는 없으나 일부 무선 기술의 레이블 처리된 데이터가 다른 무선 기술의 데이터보다 개수가 적을 때는 데이터의 불균형으로 인해 분류 성능이 낮아지는 문제가 발생하게 된다. 따라서, 데이터셋이 적게 수집된 무선 기술의 신호 데이터는 GAN을 활용해 증강시켜 전체적인 데이터의 균형을 맞추어 주게 되면 기존과 동일하게 높은 성능을 유지할 수 있다.

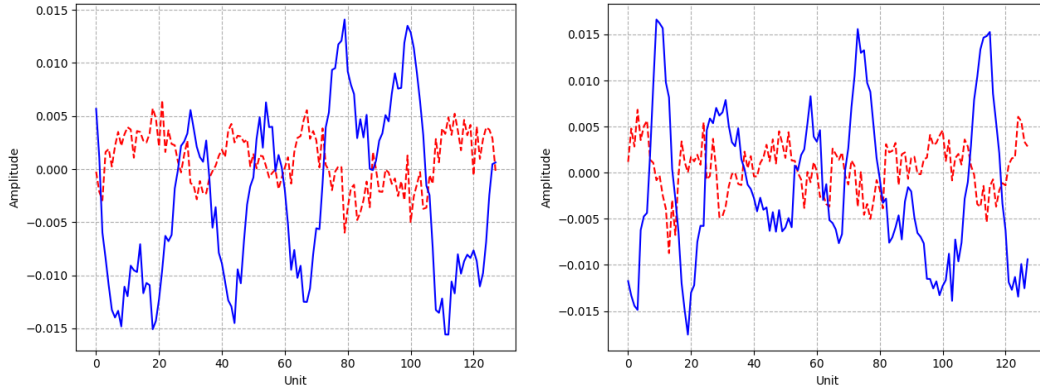
IV. 성능평가

제안하는 GAN 기반 데이터 증강 기법의 성능을 평가하기 위해 RadioML2016.10a 데이터셋을 활용하여 실험을 진행하였다. RadioML2016.10a 데이터셋은 전체 11개의 변조 기법에 대해 각각 20,000개의 신호 데이터를 포함하고 있다. 각 20,000개의 신호 데이터는 -20dB부터 18dB까지 2dB 간격으로 총 20개의 신호 대 잡음비 클래스로 나누어져 있으며, 각 클래스당 1,000개의 신호 데이터가 포함되어 있다. 전체 11개의 변조 기법 중 일반적인 환경에서 많이 적용되는 GFSK, QAM64, QPSK 기법을 대상으로 GAN 기반 데이터 증강 과정을 수행하였다.

[그림 3]은 GFSK의 실제 데이터와 GAN 기반 딥러닝 모델로 증강된 데이터의 신호의 특성을 보여준다. 원시 신호의 두 가지 입력에 대해 시간 유닛 기준으로 시각화한 결과 변조 사이의 많은 유사점을 확인할 수 있다. 이는 펄스 형성, 왜곡 및 기타 채널 효과로 인해 변조에 영향을 주는 요인들이 유사하게 적용된 신호임을 알 수 있다. 또한, 생성된 신호의 파형은 실제 신호와 매우 유사하고 약간씩 다르므로 해당 유형의 변조된 신호의 특성에 부합하고 특징의 다양성을 증가시켜 인식 네트워크의 훈련에 도움을 줄 수 있습니다.

[그림 4]는 QAM64 변조 기법으로 생성된 실제 신호 데이터와 GAN 기반 데이터 증강 기법으로 생성된 데이터를 보여준다. [그림 3]과 [그림 4]에서 보여지는 두 실험 모두 신호 대 잡음비를 8dB 수준으로 동일하게 맞추고 진행한 결과로 QAM64에 대해 생성된 신호 데이터의 경우 GFSK 변조 신호와는 다르게 진폭의 범위가 -0.0075부터 0.0075 사이에 고르게 분포하고 있어 기존 GFSK 생성 결과와 확연히 대비되는 특징을 갖는 것을 알 수 있다.

또한, [그림 5]는 QPSK 변조 기법으로 생성된 실제 신호 데이터와 GAN 기반 데이터 증강 기법으로 생성된 데이터를 보여준다. 무선 신호를 형성하는 두 입력이 -0.008 범위와 0.002 범위에서 진동하는 형태로 생성된 신호가 실제 신호의 원시 신호의 특성을 그대로 반영하고 있는 것을 알 수 있다.

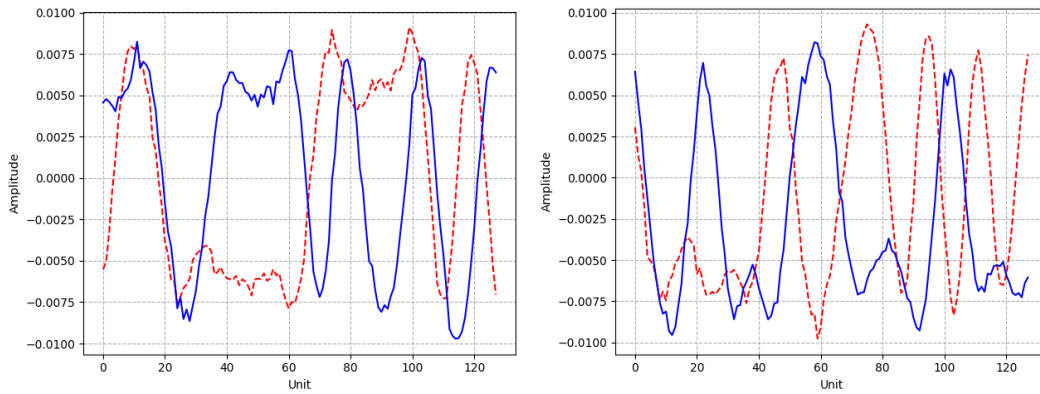


(a) Real Signal - GFSK

(b) Generated Signal - GFSK

Figure 3. Real and Generated GFSK signals under 8dB

그림 3. 8dB 의 신호 대 잡음비 환경에서의 실제 GFSK 신호와 생성된 GFSK 신호

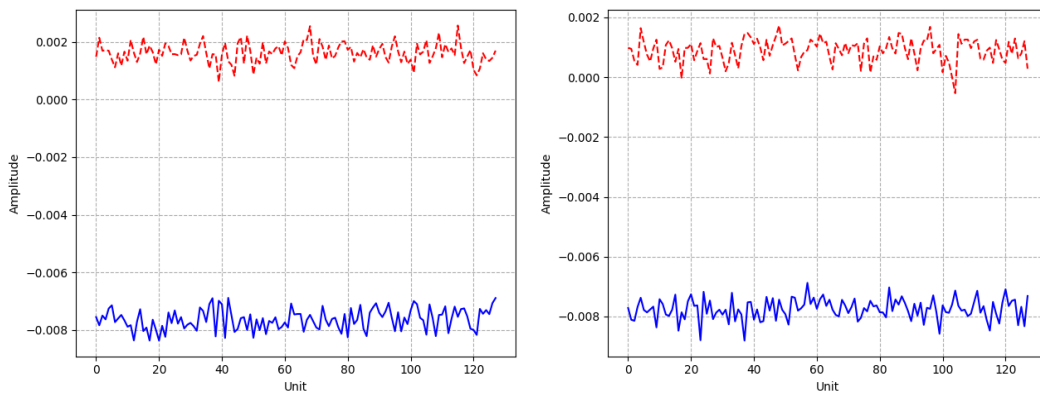


(a) Real Signal - QAM64

(b) Generated Signal - QAM64

Figure 4. Real and Generated QAM64 signals under 8dB

그림 4. 8dB 의 신호 대 잡음비 환경에서의 실제 QAM64 신호와 생성된 QAM64 신호



(a) Real Signal - QPSK

(b) Generated Signal - QPSK

Figure 5. Real and Generated QPSK signals under 8dB

그림 5. 8dB 의 신호 대 잡음비 환경에서의 실제 QPSK 신호와 생성된 QPSK 신호

본 연구에서 데이터 증강 기법과 함께 제안한 딥러닝 기반 무선 신호 분류 기법에서는 2 개의 Conv1D 레이어를 통해 원본 데이터에 대한 변환, 왜곡, 손실을 최소화하는 상황에서 데이터의 특

정을 추출하고 2 개의 LSTM 레이어를 통해 무선 신호의 시계열 특성을 고려하여 신호 분류를 수행하도록 설계하였다. 즉, 무선 신호 분석을 위해 주파수 대역 신호에 대한 원시 신호를 입력 받은 후 64 로 설정된 필터를 통과하여 데이터 판단을 위한 최소 단위로 데이터를 인식한 후 활성화 맵을 생성한다. 또한, 과대적합 문제를 고려하여 드롭 아웃을 0.6 으로 설정하여 실험을 수행하였다. 전체 데이터셋 중 각 무선 신호별로 1000 개의 샘플들을 추출하여 학습에 포함시켰다. [그림 6]에서 보는 바와 같이 100%를 모두 실제 신호로 구성하고 데이터의 균형을 맞춰 모델을 학습시킨 결과 CNN 만 사용했을 때는 약 87%의 분류 정확도를 얻을 수 있었으며, LSTM 의 경우 88%, CNN 과 LSTM 을 혼합해서 분류를 수행한 경우 92%의 높은 정확도를 얻을 수 있었다. 하지만 세 가지 GFSK, QAM64, QPSK 신호에 대해 30%의 샘플만 포함시켜 데이터 불균형 상황에서의 성능을 확인해본 결과 CNN 의 정확도는 81%, LSTM 은 80%, CNN 과 LSTM 을 혼합한 경우는 84%로 최대 8% 정도 분류 정확도가 낮아진 것을 알 수 있다. 제안한 데이터 증강 기법을 활용하여 세 가지 신호에 대해 70%의 샘플을 생성하여 실험해본 결과 CNN 의 경우 85%, LSTM 의 경우 86%, CNN 과 LSTM 을 혼합한 경우 91%로 모든 신호를 실제 데이터로 구성했을 때와 거의 유사한 결과를 얻을 수 있었다.

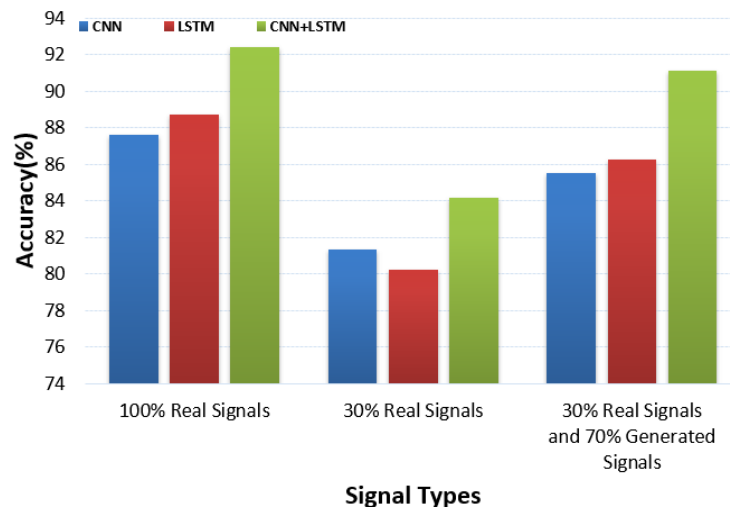


Figure 6. The Accuracy of the Wireless Signal Classification in terms of Signal Types

그림 6. 신호 구성 형태에 따른 무선 신호 분류 정확도

V. 결론

본 논문에서는 GAN 을 활용하여 무선 신호 데이터셋의 부족한 샘플을 증강시켜 신호 분류 딥러닝 모델의 성능을 향상시키는 기법을 제안하였다. 최근 딥러닝 기반 무선 신호 분류 기법들의 활용도가 높아진 상황에서 특정 신호 데이터의 부족으로 인해 데이터 불균형이 발생한 경우 전체적인 분류 정확도가 낮아지는 문제를 해결하기 위해 GAN 을 통한 데이터 증강 과정을 추가하여 딥러닝 기반 무선 신호 분류기의 전체적인 성능을 향상시킬 수 있는 것을 확인하였다. 또한, 증강된 데이터들이 기존 변조 기술의 특성에 부합하게 증강되어 전체적인 성능을 향상시킬 수 있었다. 향후 딥러닝 신호 분류 모델을 확장하기 위해 실제 데이터를 수집하는 경우 제한적인 상황에서 본 기법을 적용하여 데이터셋을 추가 확보할 수 있으며, 이를 통해 전체적인 분류 성능을 높일 수 있을 것으로 기대한다.

VI. 참고문헌

- [1] S. H. Shah and I. Yaqoob, "A survey: Internet of Things (IOT) technologies, applications and

- challenges," 2016 IEEE Smart Energy Grid Engineering (SEGE), 2016, pp. 381-385
- [2] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders and S. Pollin, "Deep Learning Models for Wireless Signal Classification With Distributed Low-Cost Spectrum Sensors," in IEEE Transactions on Cognitive Communications and Networking, vol. 4, no. 3, pp. 433-445, Sept. 2018
- [3] S. Lim, S. Lee, J. Yoo and C. Kim, "NBP: light-weight Narrow Band Protection for ZigBee and Wi-Fi coexistence", EURASIP Journal on Wireless Communications and Networking 2013, 76 (2013)
- [4] Xiaolong Zheng, Zhichao Cao, Jiliang Wang, Yuan He, and Yunhao Liu, "ZiSense: towards interference resilient duty cycling in wireless sensor networks", In Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems (SenSys '14). Association for Computing Machinery, New York, NY, USA, 119-133.
- [5] Shrahan Rayanchu, Ashish Patro, and Suman Banerjee, "Catching whales and minnows using WiFiNet: deconstructing non-WiFi interference using WiFi hardware.", In Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation (NSDI'12). USENIX Association, USA, 5.
- [6] S. Tridgell, D. Boland, P. H. W. Leong and P. H. W. Siddhartha, "Real-time automatic modulation classification," in Int. Conf. on Field-Programmable Technology, ICFPT 2019, Tianjin, China, pp. 299-302, 2019.
- [7] H. Roh, S. Oh, H. Song, J. Han and S. Lim, "Deep learning-based wireless signal classification in the iot environment," Computers, Materials & Continua, vol. 71, no.3, pp. 5717-5732, 2022.
- [8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial networks.", Commun. ACM 63, 11 (November 2020)
- [9] Alzubaidi, L., Zhang, J., Humaidi, A.J. et al, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions.", J Big Data 8, 53 (2021).
- [10] X.Wu, P.C.Y. Chen and J. Liu, "LSTMnetwork:Adeep learning approach for short-termtraffic forecast,"Iet Intelligent Transport Systems, vol. 11, no. 2, pp. 68-75, 2017.
- [11] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu et al., "Short-term residential load forecasting based on LSTM recurrent neural network," IEEE Transactions on Smart Grid, vol. 10, no. 1, pp. 841-851, 2019.
- [12] C. Weber, M. Peter and T. Felhauer, "Automatic modulation classification technique for radio monitoring," Electronics Letters, vol. 51, no. 10, pp. 794-796, 2015.
- [13] S. Zhou, Z. Yin, Z. Wu, Y. Chen, N. Zhao et al., "A robust modulation classification method using convolutional neural networks," EURASIP Journal on Advances in Signal Processing, vol. 2019, no. 1, pp. 55, 2019.
- [14] X. Zhang, J. Sun and X. Zhang, "Automatic modulation classification based on novel feature extraction algorithms," IEEE Access, vol. 8, pp. 16362-16371, 2020.
- [15] D. Hong, Z. Zhang and X. Xu, "Automatic modulation classification using recurrent neural networks," in 2017 3rd IEEE Int. Conf. on Computer and Communications (ICCC), Chengdu, China, pp. 695-700, 2017.
- [16] RadioML2016.10a Dataset, <https://www.deepsig.ai/datasets>

저자소개



임상순(Sangsoon Lim)

2013년 8월 서울대학교 컴퓨터공학과(공학박사)
 2013년 10월 ~ 2017년 8월 삼성전자 소프트웨어센터
 2017년 9월 ~ 현재 성결대학교 컴퓨터공학과 조교수

관심분야 : 인공지능, 기계학습, 모바일 컴퓨팅, 사물인터넷, 무선 네트워크