

Mel-Spectrogram 및 CNN을 활용한 응용 트래픽 분류

최정우, 김보선, 이정원, 김명섭

고려대학교, 닥터소프트

{choigoya97, boseon12}@korea.ac.kr, ljwbom@doctorsoft.co.kr, tmskim@korea.ac.kr

Classification of Application traffic using Mel-Spectrogram and CNN

Choi Jeong Woo, Kim Boseon, Lee Jeong Weon, Kim Myung-Sup

Korea Univ, Doctorsoft

요약

네트워크 관리의 기본적인 역할은 QoS (Quality of Service)를 제공하는 것이다. 적합한 QoS를 제공하기 위해서는 효과적인 응용 트래픽 분류가 필수적이다. 그러나 응용 프로그램이 다양해지고 암호화 기술의 광범위한 도입과 포트 난독화 등은 효과적인 트래픽 분류를 방해하고 있다. 이에 본 논문에서는 통계 데이터를 입력으로 사용하는 MLP 모델과 Mel-Spectrogram을 입력 이미지로 사용하는 CNN 모델을 앙상블해 응용 트래픽을 분류하는 방법을 제안한다. 제안하는 방법은 통계 데이터만을 고려한 MLP 모델만 사용한 경우와 비교하여 향상된 성능을 보였다.

I. 서론

인터넷 환경의 성장과 함께 다양한 서비스와 응용이 개발됨에 따라 트래픽이 다양해지고 있다. 이런 상황에 따라 효율적으로 네트워크를 관리하고 사용하기 위한 연구들이 진행되고 있다. 네트워크를 효율적으로 관리하는 것은 사용자에게는 고품질의 서비스를 제공하고, 운영자에게는 서비스의 신뢰성 확보로 이어질 수 있다. 네트워크 관리의 기본적인 역할은 QoS (Quality of Service)를 제공하는 것이다[1]. 적합한 QoS를 제공하기 위해 효과적인 응용 트래픽 분류가 필수적이다.

시간이 지남에 따라 트래픽 분류 기술은 많이 발전해왔다. 전통적인 방법으로는 포트 번호를 사용한 방법이나 DPI(Deep Packet Inspection)와 같은 방법들이 사용되었다. 하지만 이러한 방법들은 각각 응용 프로그램이 잘 알려진 포트 번호를 통해 트래픽을 위장하거나 일반적인 포트 번호를 사용하지 않고, 암호화되지 않은 트래픽에만 적용이 가능하다는 단점을 가진다. 따라서 RF(Random Forest)와 같은 머신러닝 모델, CNN(Convolutional Neural Network)와 같은 딥러닝 모델들이 트래픽 분류에 사용되고 있다. 하지만 머신러닝 모델의 경우 플로우 통계 정보를 사용하기 때문에 트래픽의 특성을 표현하는 것에 있어서 손실이 많다는 단점으로 인해 실용적인 응용 분류에는 적합하지 않다.

기존의 CNN 모델은 패킷 원시데이터를 일차원 또는 이차원 이미지 형태로 입력받아 분류를 수행한다. 본 논문에서는 원본 패킷 데이터로부터 헤더 정보와 통계 데이터에서 특징을 추출해 MLP의 입력으로 사용하고, 음성인식 분야에서 주로 사용되는

Mel-Spectrogram을 사용해서 이미지를 생성하여 CNN 모델의 입력으로 사용해 두 모델을 앙상블하는 응용 트래픽 분류에 대한 방법론을 제안한다. 본 논문은 2장에서 연구에 사용된 주요 기술에 대해 설명한다. 3장에서는 수집한 데이터, 데이터 전처리 과정, 분류 모델에 대해 설명하고 4장에서는 실험 결과를 설명한다. 끝으로 5장에서는 결론과 향후 연구에 대해 설명하는 순으로 진행된다.

II. 관련 연구

CNN은 시각적인 영상을 분석하는데 사용되는 다층의 피드-포워드적인 인공신경망의 한 종류이다[2]. CNN은 Feature를 추출하는 Layer와 Activation Function을 통해 결과 값을 도출하는 분류 Layer로 나뉜다. Feature 추출은 이미지의 특성을 판단하는 Convolution Layer와 추출된 Feature를 강화하는 Pooling Layer로 구성된다. 분류 Layer는 Feature 추출 Layer에서 추출된 Feature를 입력으로 받아 분류 결과를 도출한다. CNN 모델의 일반적인 구조는 그림 1과 같다.

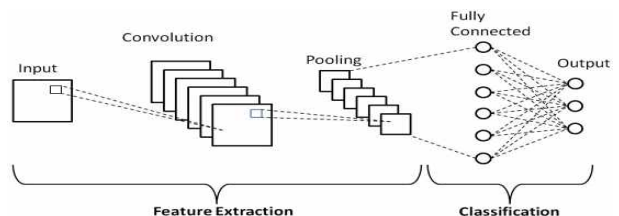


그림 1. CNN 모델 구조

Mel-Spectrogram은 음성 신호를 분석하여 시간에 대한 음성 신호의 주파수 및 강도를 나타낸 특징값이다[3]. 음성 신호를 분석하기 위해 푸리에 변환을 사용하는데, 푸리에 변환을 거친 신호는 주파수(Frequency)와 강도(Amplitude)로 분해되어 각 주파수 별 신호의 강도를 알 수 있게 된다. 시간 정보는 음성에서 매우 중요한 정보 중 하나이기 때문에, 시간 정보를 보존하면서 푸리에 변환이 가능한 단시간 푸리에 변환을 적용한다. 단시간 푸리에 변환은 신호의 처음부터 마지막까지 일정한 크기의 Window를 움직이면서 Window 내부의 신호마다 푸리에 변환을 적용하는 방법이다. 단시간 푸리에 변환의 결과로 시간, 주파수, 강도의 정보를 모두 가진 Mel-Spectrogram을 얻을 수 있다.

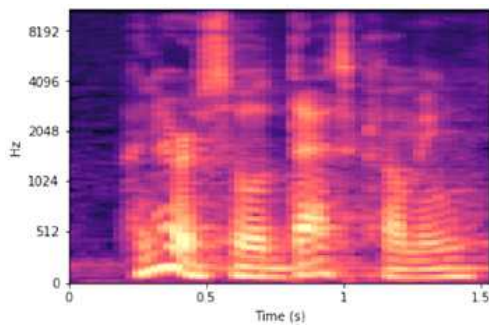


그림 2 Mel-Spectrogram 예시

III. 본론

본 장에서는 사용한 데이터셋과 데이터 전처리 과정, 분류모델에 대해 설명한다.

데이터셋은 아래의 표 1의 내용과 같이 총 10개의 응용 트래픽으로 구성되어 있다.

응용명	패킷 개수	플로우 개수
App 1	81735	901
App 2	79360	892
App 3	36481	1013
App 4	74933	936
App 5	97800	964
App 6	102963	1034
App 7	71689	1105
App 8	85215	1020
App 9	94660	881
App 10	102158	802

표 1 응용 별 패킷, 플로우 개수

MLP 모델에 사용할 특징 추출과정은 아래의 그림 3과 같다. 첫 계층에서 IAT와 SIZE로 분할되며 이후 각 특징은 패킷의 방향성에 따라 Packet Direction 단계에서 All, Forward, Backward로 나뉜다. 이후 다음 단계에서는 전체 패킷부터 시작하여 첫 5개 패킷, 첫 10개 패킷 등으로 시작하여 첫 30개까지 분할되어 추출된다. 마지막으로 Statistics 단계에서 총합, 최댓값, 척도, 왜도 등 13가지

통계 항목으로 추출된다. 최종적으로 플로우 1개당 2(Features) x 3(Direction) x 7(N) x 13 = 546 개의 특징을 추출하여 사용했다.

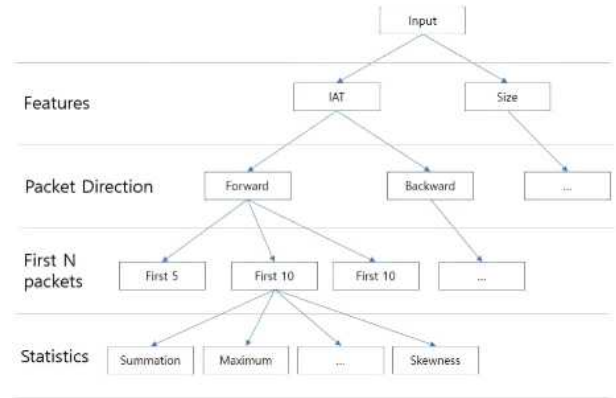


그림 3 계층적 특징 추출과정

CNN 모델의 입력으로 사용할 Mel-Spectrogram 이미지의 예시는 다음 그림 4와 같다.

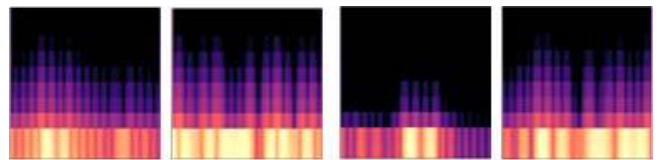


그림 4 Mel-Spectrogram 이미지

그림 4는 두 개의 응용에 대한 Mel-Spectrogram이미지이다. 맨 왼쪽부터 App 3의 처음 20개의 패킷의 패킷 사이즈, 처음 20개의 패킷의 IAT, App 7의 처음 20개의 패킷의 패킷 사이즈, 처음 20개의 패킷의 IAT에 대한 Mel-Spectrogram 이미지이다. 해당 Mel-Spectrogram 이미지의 가로축은 패킷의 개수(1~20), 세로축은 패킷 사이즈 혹은 IAT에 해당한다. 또한 해당 이미지는 패킷 사이즈, IAT와 더불어 첫 패킷부터 20번째 패킷으로 진행되면서 각 특징들의 시간에 따른 변화에 대한 정보도 표현하고 있다.

통계 데이터는 546개의 데이터를 다른 변환없이 사용하며 이때의 Shape는 1x546의 형태를 가진다. MLP 모델은 기본적인 모델을 사용했으며 3개의 Dense(뉴런 개수 64개) Layer를 가진다.



그림 5 통계 데이터 구조 및 MLP 입력 개요

Mel-Spectrogram 이미지가 입력으로 사용되는 CNN모델의 대략적인 구조는 다음 표 2와 같다.

Layer	Parameter
Input	(330, 330, 3)
Conv 2D	(n_filter=64, filter_size=(2, 2))
Pooling Layer 2D	(2, 2)
Conv 2D	(64, (2, 2))
Pooling Layer 2D	(2, 2)
Flatten	-
Dense	(10)

표 2 CNN 모델 하이퍼 파라미터

IV. 실험결과

아래의 표 3은 수집한 데이터셋에 대한 각 모델별 분류 결과를 나타낸다. 실험은 MLP 모델만 사용해 응용 트래픽을 분류한 경우와 MLP모델과 CNN모델을 Concatenate한 경우의 성능을 비교하는 방식으로 진행했다. 통계 데이터를 입력으로 사용하는 MLP 모델만으로 응용 트래픽을 분류하는 경우에 비해 Mel-Spectrogram을 입력 이미지로 사용하는 CNN 모델을 함께 사용해 응용 트래픽을 분류하는 경우가 좋은 성능을 보이는 것을 알 수 있다.

모델	Accuracy
MLP	89.56
MLP + CNN	92.78

표 3 모델 별 성능 지표

V 결론

본 논문은 통계 데이터를 입력으로 사용하는 MLP 모델과 Mel-Spectrogram을 입력 이미지로 사용하는 CNN 모델을 Concatenate하여 응용 트래픽을 분류하는 모델을 제안했으며 3%의 성능이 향상되었음을 확인했다. 향후 연구로는 응용의 종류를 50개 정도로 증가시켰을 때도 효과적으로 응용 트래픽을 분류할 수 있는 모델을 만들기 위한 연구를 수행할 계획이다.

ACKNOWLEDGMENT

본 논문은 2020년도 산업통상자원부 및 한국산업기술평가관리원(KEIT)연구비 지원에 의한 연구 (No. 20008902, IT비용 최소화를 위한 5채널 탐지기술 기반 SaaS SW Management Platform(SMP) 개발)이고, 2021년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과 (2021RIS-004)임

참고 문헌

- [1] 지세현, 백의준, 신무근, 채병민, 문호원, 김명섭, "합성곱 신경망 기반 웹 응용 트래픽 분류 모델 설계", 통신학회 논문지 Vol.44 No.05, Jun. 2019, pp. 1113-1120
- [2] KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, 25: 1097-1105.
- [3] 이정현, 윤의녕, 조근식 "전이 학습과 어텐션(Attention)을 적용한 합성곱 신경망 기반의 음성 감정 인식 모델" 2020년도 정보과학회논문 Vol.47 No.7, pp. 665-673