Ai Security

Term-Project

딥러닝 기반 네트워크 트래픽 분류

GIT-HUB 용

이름, 학번 x

Table of contents

Part 1

프 로 젝 트 개 요

목적

딥러닝 기반으로 네트워크
트래픽을 분류 하는것이 목적

이유 및 목적

네트워크 트래픽의 분류의 중요성과 사이버 보안
의 역할

정상 트래픽 네트워크와 공격 트래픽 네트워크를
딥러닝을 통해서 학습시키고 분류하는게 목표

트래픽 유형

정상적인 트래픽인 Normal
9가지 공격 유형이 존재
(Fuzzers, Analysis, Backdoors, Dos, Exploits, Generic,
Reconnaissance, Shellcode, Worms)

Part 2

Train 전처리 과정

1. 데이터 로드 및 불필요한 칼럼 제거

1. 데이터 셋을 로드하고, 분석에 필요 없는 칼럼을 제거하여 필요한 특성만 남김.

2. 데이터셋의 'id' 칼럼을 제거하여 학습에 영향을 주지 않도록 하고, attack_cat 칼럼은 타겟 변수로 사용되므로, 이를 제외한 나머지 칼럼을 학습 특성으로 사용

2. Attack_cat 인코딩 및 결측치 처리

1. 모델에서 attack_cat을 처리할 수 있도록, 숫자로 변환하고 결측치도 처리하여 학습에 영향을 주지 않도록 함.

2. attack_cat을 인코딩한 이유 : attack_cat는 문자열로 되어있어 모델이 이를 처리할 수 없기 때문에 숫자로 변환하여 모델이 이해할 수 있도록 처리

3. 범주형 변수 원핫 인코딩으로 처리

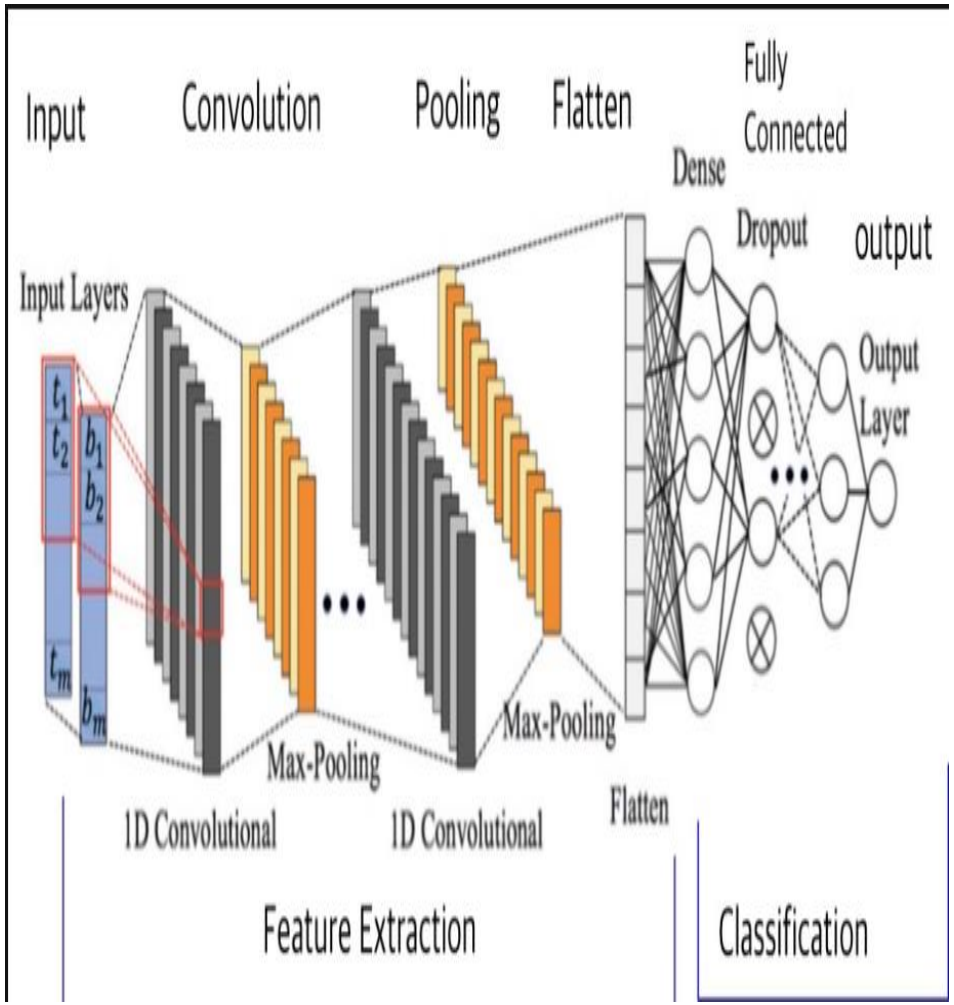
1. 범주형 데이터를 원핫 인코딩을 통해 이진 벡터로 변환하여 모델이 이를 처리할 수 있게 만듦.

2. 원핫 인코딩인 이유 : 라벨 인코딩과 달리 각 범주를 이진벡터로 변환하고 각 범주를 독립적으로 처리하여 잘못된 순서나 크기를 학습하지 않고 새로운 범주는 이를 무시하고 일관성 있게 학습을 할 수 있음

4. 데이터 스케일링 및 특성 결합

1. 변수들 간의 크기 차이를 줄여 학습 효율성을 높이고 모든 데이터를 일관된 형식으로 결합

2. 스케일링을 통해 각 특성 변수들을 정규화하여 변수 간의 크기 차이를 줄임. 원핫 인코딩으로 범주형 데이터를 기존의 수치형 데이터와 결합하여 최종 학습데이터를 만듦



CNN 모델 선정 이유

CNN 장점

1. **지역적 특징 학습**: 지역적 패턴을 인식하여 이미지나 시퀀스 데이터에서 효과적.
2. **파라미터 효율성**: 필터를 공유하여 학습할 파라미터 수가 줄어듦.
3. **공간적 구조 보존**: 이미지의 공간적 관계를 유지하며 학습 가능.

이러한 장점을 바탕으로 네트워크 트래픽 분류라는 시계열 데이터에서 일정한 패턴을 찾아내는 것이 적합하다고 생각하였고, CNN 특성상 레이어를 추가하거나 삭제하는 하는 등 모델을 확장에도 용이하여 선택.

Training

```
Epoch 1/10, Train Loss: 0.6805, Val Loss: 0.5691, Val Acc: 78.07%
Epoch 2/10, Train Loss: 0.5530, Val Loss: 0.5336, Val Acc: 79.04%
Epoch 3/10, Train Loss: 0.5278, Val Loss: 0.5246, Val Acc: 78.99%
Epoch 4/10, Train Loss: 0.5117, Val Loss: 0.5054, Val Acc: 80.19%
Epoch 5/10, Train Loss: 0.5008, Val Loss: 0.5053, Val Acc: 80.01%
Epoch 6/10, Train Loss: 0.4919, Val Loss: 0.4925, Val Acc: 80.60%
Epoch 7/10, Train Loss: 0.4850, Val Loss: 0.4910, Val Acc: 80.46%
Epoch 8/10, Train Loss: 0.4795, Val Loss: 0.4832, Val Acc: 80.89%
Epoch 9/10, Train Loss: 0.4756, Val Loss: 0.4818, Val Acc: 80.91%
Epoch 10/10, Train Loss: 0.4733, Val Loss: 0.4808, Val Acc: 81.00%
Model saved to /content/drive/MyDrive/train_save/cnn_model_epoch.pth
```

학습 과정 및 정확도

모델 설정 : 10 Epoch, 학습율 조절 : 1e-4, 배치 사이즈 : 32로 설정

1. Train Loss 와 Val Loss

Train Loss : 0.6805 -> 0.4733 로 감소

Val Loss : 0.5691 -> 0.4808 로 감소

2. 검증 정확도

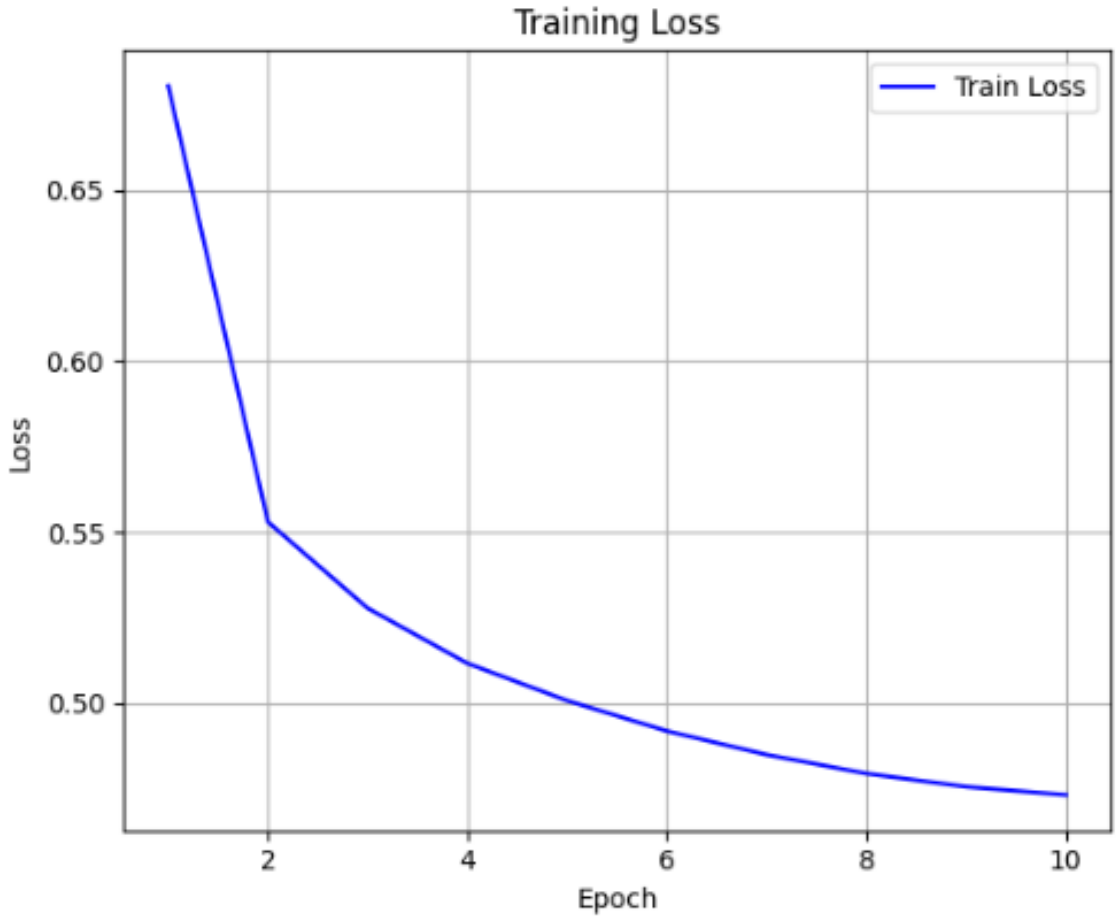
Epoch : 1 일 때, 78.07% -> Epoch : 10 일 때 81%로 증가

정확도는 점진적으로 상승

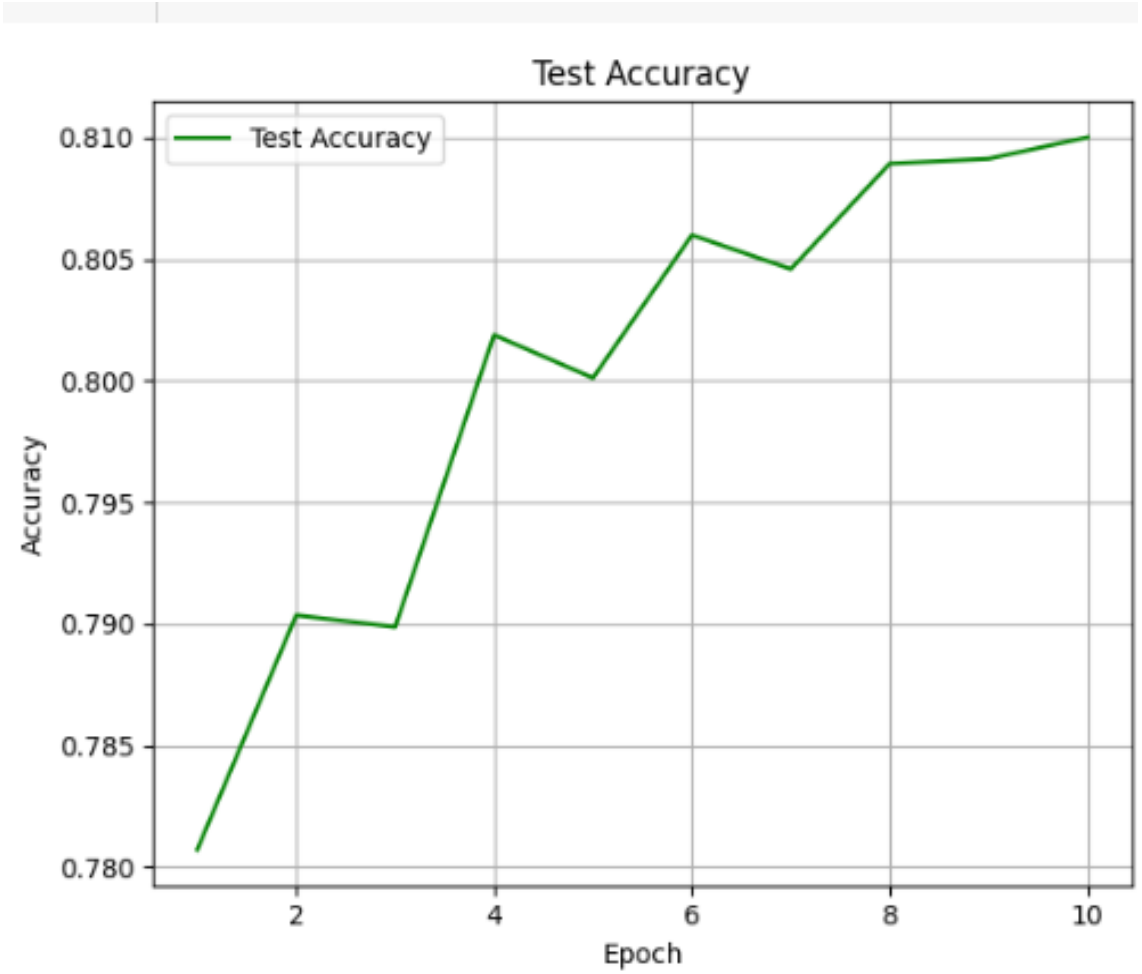
이를 통해 학습을 통해서 손실율은 감소,
검증 정확도는 증가 했다는 것을 알 수 있음.

모 델 학 습 및 검 증

Training Loss



Training Val Accuracy



Part 6

Test 전처리 과정

1. Id 컬럼 처리

1. Id 컬럼에서 데이터에서 분리하여, 해당 정보를 별도로 저장하고 데이터를 제거하여 모델 학습에 방해하지 않도록 함
2. Id 컬럼 제거한 이유 : 모델 학습에 영향을 미치지 않기에 학습에 방해되지 않도록 제거함.

2. 범주형 칼럼 처리

1. Train code와 마찬가지로 변수들을 모델이 처리할 수 있는 형태로 변환
2. 가능한 범주형 변수들을 원핫 인코딩 함.
3. 원핫 인코딩은 미리 train_code에서 저장한 pkl를 로드하여 수행.

3. 데이터 변환 및 스케일링

1. 데이터를 모델에 적합한 형태로 변환하고, 각 특성 간의 차이를 줄여 학습의 효율성을 높임
2. 원핫 인코딩을 통해 인코딩 된 데이터와 기존의 수치형 데이터가 결합된 형태로 변환
3. Trian code에서 저장한 scaler을 로드하여 스케일링 수행

4. 반환 값 처리

1. 전처리된 데이터와 필요시 id 정보 반환
2. Return_ids가 참이면 id와 함께 전치리된 데이터를 반환
3. Return_ids가 거짓이면 id 없이 전처리된 데이터만 반환

Part 7

테 스 트 결 과

테스트 샘플 수 확인

```
attack_cat counts:
Normal: 27750 samples
Fuzzers: 12236 samples
Analysis: 373 samples
Backdoor: 1 samples
DoS: 422 samples
Exploits: 18851 samples
Generic: 18045 samples
Reconnaissance: 4164 samples
Shellcode: 490 samples
Worms: 0 samples
```

테스트 정확도

```
Test Accuracy: 75.05%
```

테스트 예측 결과

Train code에서 정확도가 81%라는 준수한 성능이 나와,
Test code에서도 그와 비슷한 성능이 나올 거라고 예측.

예측한 칼럼과 실제 칼럼을 비교 결과 75.05%라는 결과가 나옴
하지만 Worms 공격 유형에 대한 값이 0 이라 예측일 잘 하지 못 했다는 걸 알 수 있음

정확도는 높은 편이지만, 특정 공격 유형에 대한 예측은 실패 함.

이를 통해, 데이터 불균형 문제나 특정 공격 유형에 대한 추가 학습이 필요하다는 걸 알 수 있음.

Part 8

참 고 문 헌

<Git hub>

- https://github.com/2hyes/security_ml
- https://github.com/Hashehri/Network-Traffic-Classification-UNSW-NB15/blob/main/code/ML/testmodeling.ipynb

<사이트>

- https://research.unsw.edu.au/projects/unsw-nb15-dataset <UNSW 대학 NB15 데이터셋>

<논문>

- https://arxiv.org/abs/2106.12693 - <deep learning for network taffic classification>
- https://scienceon.kisti.re.kr/srch/selectPORSrchArticle.do?cn=JAKO201935236776145

<네트워크 공격 탐지 성능향상을 위한 딥러닝을 이용한 트래픽 데이터 생성 연구>

- https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART002935940

<마이터 어택과 머신러닝을 이용한 UNSW-NB15 데이터셋 기반 유해 트래픽 분류>

<기타>

- [9주차]_AI보안_실습_NIDS 실습 소스.


감사합니다