유해 네트워크 트래픽 탐지를 위한 컨볼루션 신경망기반 트래픽 분류 기법 연구

염성웅, 뉘엔 지앙 쓰엉, 뉘엔 반 퀴엣, 김경백
전남대학교 전자컴퓨터공학부
yeom4032yeom4032@gmail.com, truongnguyengiang.bk@gmail.com,
quyetict@gmail.com, kyungbaekkim@jnu.ac.kr

A Study on Convolutional Neural Network based Traffic Classification Methods for Detecting Malicious Network Traffic

Sungwoong Yeom, Giang-Truong Nguyen, Van-Quyet Nguyen, Kyungbaek Kim
Dept. Electronics and Computer Engineering, Chonnam National University

요 약

최근 유해 네트워크 트래픽을 탐지하기 위해 머신러닝 기법을 활용하는 방법론이 주목을 받고 있다. 이 논문에서는 딥러닝 기법중 하나인 컨볼루션 신경망 (Convolutioanl Nural Network)을 기반으로 유해 네트워크 트래픽을 분류하는 기법을 소개한다. 우선 이미지 처리에 강한 컨볼루션 신경망을 활용하기 위해, 네트워크 트래픽의 주요 정보를 규격화된 이미지로 변환하는 방법을 제안한다. 이후 네트워크 트래픽 정보를 변환한 이미지를 입력으로 컨볼루션 신경망을 학습을 시켜 제공되는 네트워크 트래픽의 분류를 수행하도록 한다. KDD 1999 데이터셋을 활용하여 이미지 변환 및 컨볼루션 신경망 기반 네트워크 트래픽 분류 기법의 성능을 검증하였다. 특히, 이미지 변환에 이용되는 트래픽 정보의 변동에 대해서 컨볼루션 신경망 기반 네트워크 트래픽 분류기법이 안정적으로 동작하는 것을 확인하였다.

Ⅰ. 서 론

네트워크 기술의 발달과 IoT 기기의 활성화에 따른 네트워크 상의 트래픽의 복잡도가 점점 높아지면서, 네트워크에 유해 트래픽을 유발시켜서 네트워크 서비스의 질을 저하시키거나 특정 서버 및 호스트의 동작에 피해를 입히는 네트워크 공격에 대한 탐지 및 방어가 더욱 중요해지고 있다. 최근, 머신러닝 기반의 네트워크 공격 트래픽 분류기법에 대한 연구가 주목을 받고 있다.[1][2] 이 연구들에서는 주로 종단간 연결정보, 도메인정보, 데이터전송정보와 같은 여러 네트워크 트래픽 정보를 특징벡터로 이용하는 다양한 분류기 (SVM, KNN, Naive Bayes)를 활용하여 네트워크 공격 트래픽을 분류하는 방법을 제안하였다.

본 논문에서는 딥러닝 기법중 하나인 컨볼루션 신경망 (Convolutional Neural Network)를 이용하여 네트워크 공격 트래픽을 분류하는 기법을 소개한다. 제안하는 기법은 KDD 1999 데이터셋에서 제공되는 네트워크 트래픽 정보를 규격화된 이미지로 변환하고, 변환된 이미지들을 이용해 컨볼루션 신경망 모델을 학습시켜 향후 네트워크 트래픽 분류를 위한 모델을 도출한다. 학습된 컨볼루션 신경말 모델을 이용해 입력되는 트래픽 정보가 일반적인 네트워크 트래픽인지 네트워크 공격에 사용되는 트래픽인지를 구분한다. KDD 1999 데이터 셋을 활용한 검증을 통해 제안되는 기법이 기존의 머신러닝 기반의 방법보다 성능이 우수함을 확인하였고, 또한 컨볼루션 신경망에 활용되는 이미지 변환 시, 특징벡터의 순서 및 그룹화가 성능에 미치는 영향이 미미함을 확인하였다.

2장에서는 관련 머신러닝 기법 및 컨볼루션 신경망에 대해 소개하고, 3장에서는 제안하는 컨볼루션 신경망 기반 유해 네트워크 트래픽 탐지 기법을 소개한다. 4장에서는 KDD 1999 데이터 셋에 기반한 제안 기법의 성능 검증 결과를 기술하고, 5장에서 본 논문의 결론 및 향후 연구 내용에 대해 기술한다.

Ⅱ. 관련연구

1. SVM (Support Vector Machine)

SVM은 주어진 특징 벡터들 간의 마진을 최대로 하는 방법을 이용해 서로 다른 특징을 가지는 데이터 집합을 분류하는 기법으로, 다수의 특징 벡터가 주어지더라도 집합간의 Support Vector를 구함으로써 분류기법을 안정적으로 운용할 수 있는 장점을 가진다.

2. KNN (K-Nearest Neighbors)

KNN은 임의의 데이터를 입력으로 이용하였을 때, 해당 데이터의 특징 벡터와 다른 데이터들의 특징 벡터와 유사도를 계산하여, 가장 유사도가 높은 K개의 데이터를 이웃으로 선택하는 기법이다. 만약 K = 1이고, KNN을 분류 기법으로 이용한다면, 입력된 데이터는 가장 유사도가 높은 하나의 그룹으로 분류된다.

3. Naive Bayes

Naive Bayes 분류 기법은 Bayes 이론을 적용하는 확률적 분류기법으로, 특징 벡터들 간의 독립성이 강할수록 그 성능이 좋아진다.

4. 컨볼루션 신경망 (Convolution Neural Network)

컨볼루션 신경망 (CNN : Convolution Neural Network) 는 영상 이미지 분류를 위한 최신의 분류 모델로, 다수의 필터를 영상 이미지의 픽셀 데이터에 적용하여 고차원 특징을 추출하여 분류기를 학습하는 모델이다. 이때, 추출되는 고차원 특징들은 convolution layer, pooling layers, fully connected layer로 구성되는 hidden layer 내부에 존재하게 되어 각 특징들에 대한 자세한 정보를 확인하기 어렵고, 특별한 의미를 부여하기 힘들다. Convolution layer는 영상 이미지의 여러 sub-region에 다양한 convolution filter를 적용하여 여러 입력을 하나의 출력으로 계산하는 비선형적 수학적 계산 모델을 적용한 계층이다. Pooling layer는

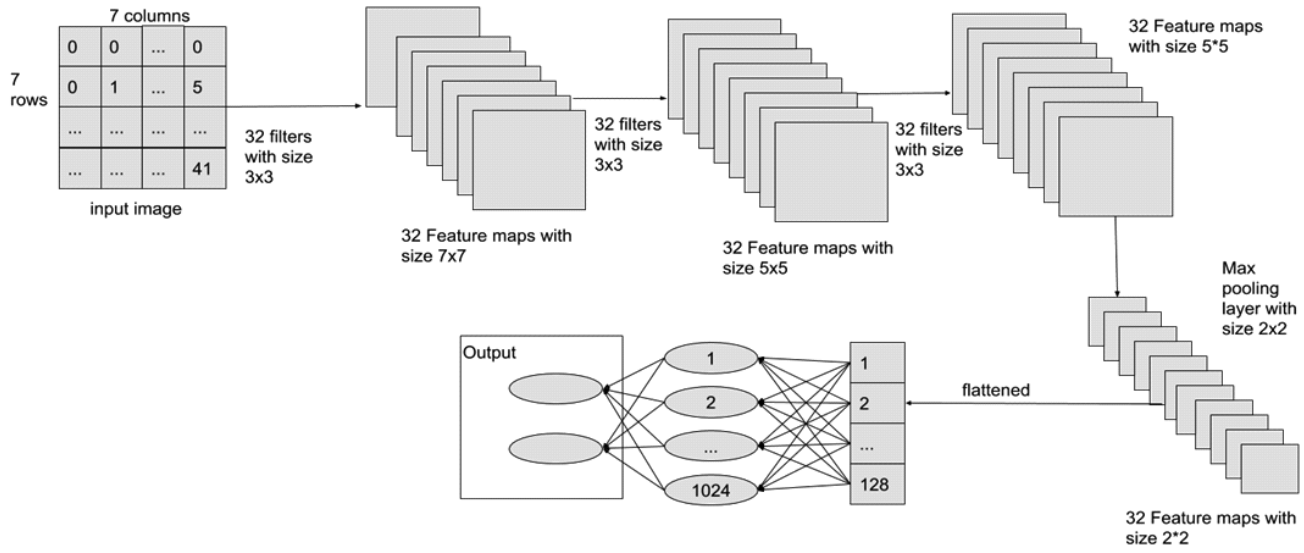

그림 2. 제안된 컨볼루션 신경망 기반 네트워크 트래픽 분류 모델 구조

convolution layer에서 계산된 데이터 계층의 크기를 줄이는 계층이고, fully connected layer는 입력되는 데이터 계층의 노드들의 값을 입력으로 사용하고 추가적인 hidden 노드들을 활용하여 출력 값을 계산하는 신경망 모델로, Supervised 또는 Unsupervised learning을 통해 해당 신경망을 학습시킨다.

Ⅲ. 컨볼루션 신경망 기반 유해 네트워크 트래픽 탐지 기법

본 논문에서는 네트워크 트래픽 정보를 이미지로 표현하고, 이를 컨볼루션 신경망을 통해 학습시켜 네트워크 공격 트래픽을 분류하는 모델을 제공하는 기법을 제안한다.

1. 네트워크 트래픽 정보의 이미지 변환

제안하는 기법을 위해 우선적으로 네트워크 트래픽 정보를 이미지로 변환하는 것이 필요하다. 우리는 KDD 1999 데이터 셋을 기준으로 네트워크 트래픽 정보를 표현하는 특징 벡터를 추출하였다. 사용되는 특징벡터는 총 41가지로 TCP connection 특징 9가지, Domain knowledge 관련 connection 특징 13가지, 2초간의 connection traffic 특징 9가지, 그리고 2초 이상에 해당하는 공격 특징 10가지를 포함한다.[3] 네트워크 트래픽 특징벡터를 이미지로 변환하기 위해, 모든 특징 백터 값을 0부터 255사이의 값으로 Normalize하고, 가로 7 픽셀, 세로 7픽셀을 가지는 정사각형 이미지로 변환하였다.

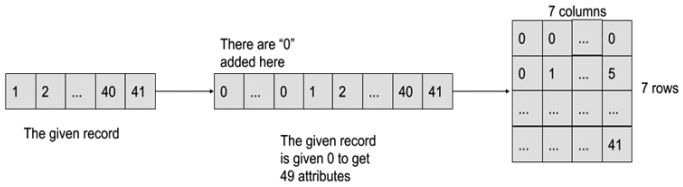

그림 1. 네트워크 트래픽 특징 벡터의 이미지 변환

특징벡터를 7X7 이미지로 변환할 때, 어떤 벡터를 이미지의 몇 번째 셀에 맵핑할지는 두가지 방법론을 사용하여 수행하였다. 첫 번째는, KDD 데이터 셋에서 제공되는 데이터 순서를 지키고 데이터 앞쪽에 0 값을 더하여 변환을 수행하는 방법이다. 두 번째 방법은 앞서 말했던 특징 그룹별로 데이터 셋을 재 정렬한 후 데이터 중간 중간에 0을 삽입하여 이미지로 변환하는 방법이다. 이러한 두 경우에 대해서 실험을 수행한 결과 공격 트래픽 탐지 결과가 유사하게 나오는 것을 확인하였다.

2. 네트워크 트래픽 분류를 위한 컨볼루션 신경망 구조

제안하는 방법에서 사용한 컨볼루션 신경망 구조는 그림 2와 같다. 7 X 7 크기의 이미지를 입력으로 사용하고, 총 3개의 convolution layer를 사용하고, 1번의 pooling을 수행 후, fully connected layer를 학습한다. Convolution layer의 경우, 3 X 3 크기의 32개 필터를 적용하여 layer를 생성하는데, 두 번째 레이어를 생성할 때는 padding을 하지 않고 convolution layer의 크기를 5 X 5로 줄인다. Pooling layer에서는 총 2x2x32 = 128개의 값을 노드로 가지게 되고, fully connected layer에서 1024개의 hidden node를 학습시킨다. 본 논문에서는 네트워크 트래픽을 일반 트래픽과 공격 트래픽 (DoS)을 구분하는 분별기를 학습시켜서 출력이 2개로 나오도록 구조를 설계하였다.

Ⅳ. 검증

제안된 기법의 성능을 검증하기 위해, KDD 1999 데이터 셋을 기반으로 Cross Validation을 수행하여, 각 분류 기법별 Accuracy와 False positive를 측정하였다. 검증을 위해 사용된 데이터 셋은 KDD 1999데이터 중 일반 네트워크 트래픽과 DoS공격 네트워크 트래픽에 대한 데이터를 선별하여 준비하였다. 전체 데이터의 90%는 분류기 모델 학습에 사용하였고, 10%는 모델 성능 테스트로 사용하였고, 10-fold Cross Validation을 수행하였다.

비교하는 분류모델로는 SVM, KNN, Naive Bayes 그리고 제안하는 CNN기반 분류모델이 있다. SVM, KNN, Naive Bayes 모델은 Weka 3.8.1을 활용하였다. [4] 제안하는 CNN기반 분류모델은 TensorFlow를 활용해 구현하였다.

그림 3에서는 네트워크 트래픽 분류기법 별 Accuracy (true detection / true attack traffic)를 나타낸다. 그림 4에서는 네트워크 트래픽 분류 기법별 False Positive (false detection/true normal traffic)를 나타낸다. 이 결과에서 Naive Bayes가 가장 성능이 좋지 않은 것을 확인할 수 있고, 제안

되는 CNN 기반 모델이 가장 성능이 좋은 것을 확인할 수 있다. 특히 CNN 기반 모델은 92%이상의 Accuracy를 달성하면서 2%이하의 False positive를 가지는 것을 확인할 수 있었다.

또한, CNN 기반 네트워크 트래픽 분류 기법은 네트워크 트래픽의 특징 벡터를 이미지로 바꾸는 매핑 방법에 무관하게 일정하게 높은 성능을 유지하는 것을 확인할 수 있었다.

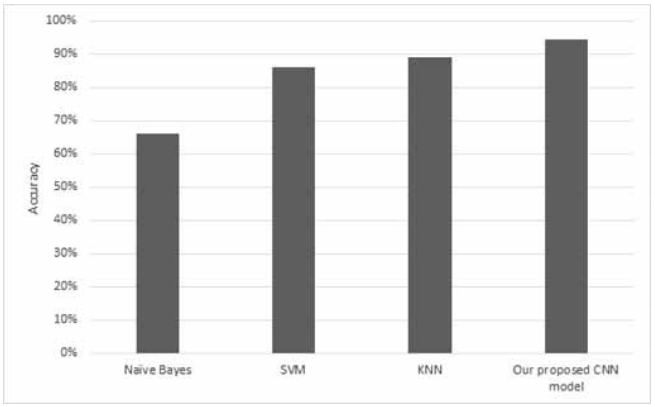

그림 3. 네트워크 트래픽 분류기법 별 Accuracy

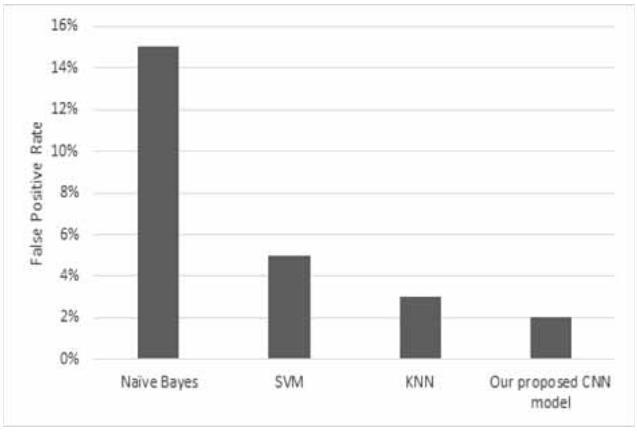

그림 4.네트워크 트래픽 분류기법 별 False Positive

V. 결론

본 논문에서는 유해 네트워크 트래픽을 탐지하기 위한 컨볼루션 신경망(CNN) 기반 네트워크 트래픽 분류 기법을 제안하였다. 제안하는 기법은 네트워크 트래픽의 특징 벡터를 이미지로 변환하고 이를 CNN을 적용하여 분류기 학습에 이용함으로써, DoS 공격 네트워크 트래픽과 일반 네트워크 트래픽을 성공적으로 분류한다. 특히, 특징 벡터의 이미지 변환 시 사용되는 매핑 방법이 분류기 성능에 영향이 거의 없음을 확인하였다. 즉, 네트워크 분류를 위해 원하는 특징벡터를 보다 유연하게 활용할 수 있다는 점을 확인하였다.

제안된 기법은 KDD 1999 데이터셋에서 제공되는 트래픽 특징 벡터를 기반으로 분류 모델을 학습하였는데, 현재 제안된 기법을 적용하여 실제 네트워크 시스템에서 실시간으로 해당 모델을 학습시키는 방법에 대한 연구를 진행 중이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017RIA2B4012559)). 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터 지원사업의 연구결과로 수행되었음 (IITP-2019-2016-0-00314).

참 고 문 헌

[1] 최진태, 뉘엔 신 응억, 김경백, “유해 네트워크 트래픽 탐지를 위한 트래픽 분류 기법 성능 비교”, 2017년도 한국인터넷정보학회 추계학술대회 논문집 제 19권 2호, 2017

[2] Jintae Choi, Sinh-Ngoc Nguyen, Jeongnyeo Kim, Guee-Sang Lee, Kyungbaek Kim, “Performance Comparison of Traffic Classification Techniques for Detecting Malicious Network Traffic”, In the proceedings of SMA 2017 conference, December 2017.

[3] “KDD Cup 1999 Data.” [Online]. Available: http://kdd.ics.uci.edu/database/kddcup99/kddcup99.html, November 2017.

[4] “Waikato environment for knowledge analysis (weka) version 3.8.1.” [Online] Available : http://www.cs.waikato.ac.nz/ml/weka, October 2017.