플로우 간 관계성을 고려한 딥러닝 기반 응용 트래픽 분류

백의준, 김보선, 이민성, 김명섭

고려대학교

{pb1069, boseon12, min0764, tmskim}@korea.ac.kr

DL-based Application Traffic Classification Considering Relationship between Flows

Ui-jun Baek, Boseon Kim, Min-Seong Lee, Myung-Sup Kim

Korea University

요 약

인터넷 환경의 급격한 성장에 따라 서비스 품질 관리, 자용 사용 정책 수립, 침입 탐지 등을 위한 네트워크 응용 트래픽 분류의 중요성이 높아지고 있다. 그러나, 응용 프로그램의 다양성과 암호화 기술의 광범위한 도입 그리고 포트 난독화 등은 실용적인 트래픽 분류를 방해하고 있으며 이에 새로운 플로우 통계 특징과 최신 딥러닝 모델을 활용한 많은 연구들이 수행되고 있으나 실제 환경에서 활용하기 어렵다. 이에 본 논문에서는 응용 트래픽 내 플로우 관계성을 고려한 앙상블 기반 분류 방법을 제안한다. 제안하는 방법은 단일 트래픽의 특징만을 고려한 분류 방법과 비교하여 향상된 성능을 보였다.

Ⅰ. 서 론

인터넷 환경의 급격한 성장에 따라 서비스 품질 관리, 자원 사용 정책 수립, 침입 탐지를 위한 네트워크 응용 트래픽 분류의 중요성이 높아지고 있다. 그러나 응용 프로그램의 다양성과 암호화 기술의 광범위한 도입 그리고 포트 번호 난독화 등은 실용적인 트래픽 분류 기술 도입을 방해하고 있으며 이에 새로운 플로우 통계 특징과 시계열 특성을 반영하는 딥러닝 기반 분류 모델에 대하여 활발히 연구되고 있다.

트래픽 분류 기술은 시간이 지남에 따라 크게 발전해왔다. 첫 번째로 가장 전통적이며 쉬운 방법인 포트 번호을 사용한 분류 방법이 있으며 최신 응용프로그램이 잘 알려진 포트 번호를 통해 트래픽을 위장하거나 일반적인 포트 번호를 사용하지 않기 때문에 정확도가 떨어진다. 또 하나의 전통적인 방법으로는 DPI(Deep Packet Inspection)가 있으며 패킷 내 페이로드에서 응용을 분류할 수 있는 특정 패턴 혹은 키워드를 찾는데 중점을 두지만 암호화되지 않은 트래픽에만 사용할 수 있다는 단점을 가진다. 세 번째로는 플로우 통계 특징을 사용하는 분류 방법이 있으며 암호화 여부에 상관없이 분류를 수행할 수 있다는 장점을 가지고 주로 RF(Random Forest)와 같은 머신러닝 모델이 사용되었다. 그러나 인위적으로 추출한 특징이 트래픽의 특성을 표현함에 있어 많은 손실이 있기에 간단한 응용 분류에서는 높은 성능을 보이지만 실용적으로 사용하기는 어렵다. 마지막으로 딥러닝 기반 분류가 있으며 주로 패킷 원시데이터를 일차원 또는 이차원 이미지 형태로 입력받아 분류를 수행하는 CNN(Convolutional Neural Network)과 RNN(Recurrent Neural Network) 또는 이들의 합성 모델이 주로 사용되고 있다. 이러한 활발한 연구에도 불구하고 수십 개부터 수백 개까지 응용 프로그램이 수행되는 실제 환경에 도입하기에는 무리가 있다. 이에 본 논문에서는 응용 트래픽 내 플로우 관계성을 고려한 앙상블 분류 방법을 제안한다. 제안하는 방법은 단일 트래픽 플로우에 대한 분류 결과와 해당 플로우와 이미 분류된 플로우와의 관계 점수를 동시에 고려하여 분류 성능을 높였다.

본 논문은 서론에 이어 2장에서 사용된 주요 기술에 대하여 설명한다. 3장에서는 수집한 데이터와 분류 모델에 대해 설명하고 4장에서는 실험 결과를 설명한다. 마지막으로 5장에서는 결론 및 향후연구에 대해 설명하고 논문을 마친다.

Ⅱ. 관련 연구

CNN은 시각적 영상을 분석하는 데 사용되는 다층의 피드-포워드적인 인공신경망의 한 종류이다[1]. CNN은 크게 특징을 추출하는 Layer와 활성함수를 통해 결과 값을 도출하는 분류 Layer로 나뉜다. 특징 추출은 이미지의 특징을 판단하는 Convolutional Layer와 추출된 특징을 강화하는 Pooling Layer로 구성된다. 분류 Layer는 특징 추출 Layer에서 추출된 특징을 입력받아 분류 결과를 도출하며 일반적인 구조는 그림 1과 같다.

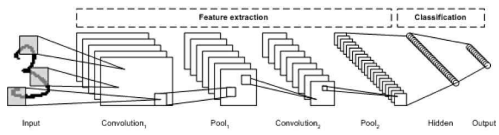

그림 1. CNN 구조

RNN[2]은 입력 데이터의 시계열 특성을 고려하는 인공 신경망의 한 종류로 유닛 간 연결이 순환적 구조를 가진다. RNN은 은닉층의 노드에서 활성 함수를 통해 결과 값을 출력층으로 내보냄과 동시에 다음 은닉층의 입력으로 전달하는 특징을 가지며 일반적인 구조는 그림 2와 같다.

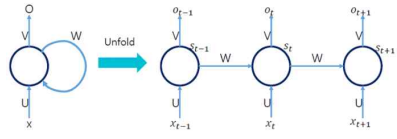

그림 2. RNN 구조

이 연구는 2020년도 산업통상자원부 및 한국산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임 (No. 20008902, IT비용 최소화를 위한 5채널 탐지기술 기반 SaaS SW Management Platform(SMP) 개발)

Ⅲ. 분류

본 장에서는 사용한 데이터셋과 분류 모델에 대하여 설명한다.

데이터셋은 각기 다른 응용 트래픽 7종으로 구성되어 있으며 자세한 내용은 표 1과 같다.

App No.	Flow		packet	
	#	%	#	%
App1	448	62%	29366	61%
App2	73	10%	2337	5%
App3	65	9%	1673	3%
App4	27	3%	3261	7%
App5	57	7%	6082	13%
App6	50	7%	4809	10%

표 1. 응용 별 플로우 개수 및 비율

본 연구에서는 분류 모델로 [3]에서 사용한 모델 구조를 차용한 경량모델을 사용했으며 모델 후반부에 배치되는 시계열 모델로 LSTM(Long Short Term Memory)이 아닌 GRU(Gated Recurrent Unit)를 채택하였다. 단일 플로우 분류 모델과 플로우 관계 검사모델 모두 입력 레이어의 개수를 제외하곤 동일한 하이퍼 파라미터를 적용했다. 모델의 하이퍼 파라미터는 표2와 같으며 전체 분류 모델 구조는 그림 3과 같다.

Layer	Type	Filters/Neurons
1	conv2D+relu	2
2	max pooling 2D	2
4	dense	1
5	gru	10
6	dense	7

표 2. 분류 모델의 하이퍼 파라미터

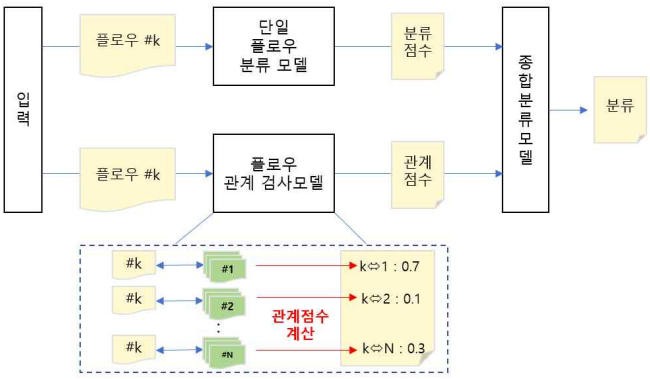

그림 3. 분류 모델 구조

제안하는 분류 모델은 크게 단일 플로우 분류 모델과 플로우 관계 검사모델로 나뉜다. 단일 플로우 분류 모델은 단일 플로우가 해당 플로우가 어떤 응용에 속하는지 판단하고 각 응용에 속할 확률을 도출한다. 플로우 관계 검사 모델은 입력되는 플로우와 기존에 분류된 플로우와의 관계 점수를 계산하며 관계 점수는 입력된 플로우와 한 클래스 내 속한 모든 플로우와의 관계 점수의 평균으로 계산한다. 단일 플로우 분류 점수와 플로우 관계 점수는 종합분류모델에 입력되어 최종 분류 결과를 도출한다. $SFCR_{k,i}$가 입력 플로우 k와 클래스 i에 대한 단일 플로우 분류 점수이고 $RFS_{k,i}$가 입력 플로우 k클래스 i에 대한 플로우 관계 점수일 때 종합 분류 점수 S에 대한 수식은 다음과 같다:

$$S_{k,i} = \frac{SFCR_{k,i} \times (1-\alpha) + RFS_{k,i} \times \alpha}{2} \quad (1)$$

이때, α는 단일 플로우 분류 점수와 플로우 관계 점수의 비율을 조정하는 매개변수이다.

Ⅳ. 실험 결과

표 3은 수집한 데이터셋에 대한 각 모델별 분류 결과를 나타낸다. 제안한 모델은 [3]보다 모든 평가지표에서 좋은 성능을 보이는 것을 알 수 있으며 특히 FPR의 경우 반이 줄어들었으며 Recall의 경우 9%, Accuracy의 경우 6퍼센트 가량 좋은 것을 확인할 수 있다. 이 외 단순한 CNN 모델들은 낮은 성능을 보이는 것을 확인했다.

모델	Accuracy	Recall	FPR
1d CNN	78.26	82.60	26.08
2d CNN	68.26	62.39	25.86
3d CNN	84.45	79.78	10.87
[3]	92.17	89.56	5.28
proposed	98.03	98.68	2.6

표 3. 모델 별 성능지표

그림 4는 매개변수 α의 변화에 따른 분류정확도를 나타낸다. 실험 결과 α의 값이 0.4 ~ 0.6에 위치할 때 좋은 성능을 보이는 것을 확인했으며 이는 단일 플로우 탐지 점수만큼이나 새로 제안한 플로우 관계에 대한 특징이 유의미하다는 것을 의미한다.

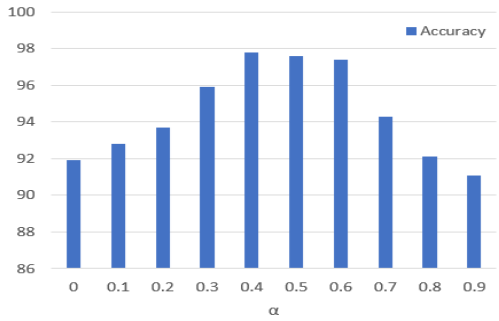

그림 4. 매개변수α에 따른 분류 정확도

Ⅴ. 결론

본 논문은 응용 트래픽 플로우 간 관계성을 고려하여 분류하는 모델을 제안하였으며 기존 방법 대비 10%의 성능 향상을 이끌어 내었다. 그러나 본 방법은 초기 분류되는 결과에 따라 이후 분류 결과가 악화될 수 있다는 한계점을 가진다. 따라서 우리는 초기 분류 결과에 따른 편향성을 줄이기 위한 연구를 수행할 계획이다.

참 고 문 헌

[1] KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, 25: 1097-1105.

[2] RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning representations by back-propagating errors. nature, 1986, 323.6088: 533-536.

[3] WANG, Wei, et al. HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection. IEEE access, 2017, 6: 1792-1806.