Tensorflow Softmax Regression을 적용한 웹 브라우저 응용 트래픽 분류 모델 생성

박지태, 심규석, 이성호, 김명섭

고려대학교

{pjj5846, kusuk007, gaek5, tmskim} @korea.ac.kr

Web Browser Application Traffic Classification Models Using Tensorflow Softmax Regression

Jee-Tae Park, Kyu-Seok Shim, Sung-Ho Lee, Myung-Sup Kim

Korea Univ.

요 약

네트워크 환경이 날이 갈수록 비약적으로 발전함에 따라 응용(Application)은 복잡해지고 다양해지고 있다. 이에 응용 트래픽을 정확하게 분류 하는 방법의 중요성도 같이 커지는 추세이다. 응용 트래픽을 분류하는 방법은 다양하게 있지만 요즘에는 기계학습을 이용한 방법이 새롭게 대두하고 있다. 이는 기계학습이 날이 갈수록 패턴이 다양하고 복잡해지는 트래픽을 정확하고 효율적으로 분류하는데 기존의 페이로드 시그니처 방법에 비해 적합하기 때문이다. 따라서 본 논문에서는 기계학습 기반의 응용 트래픽 분류 하는 방법을 제안한다. 제안하는 방법을 통해 설계된 응용 트래픽 분류 모델을 웹 브라우저 응용 트래픽에 적용했을 때 약 95%의 Accuracy를 얻을 수 있었다.

Ⅰ. 서 론

오늘날 네트워크 환경은 점점 복잡해지고 커지고 있고, 네트워크 응용 트래픽 종류도 증류와 패턴도 다양해지고 복잡해지고 있다. 이러한 추세에 따라 응용 트래픽을 정확하게 분류 하는 방법의 중요성도 같이 커지고 있다. 응용 트래픽을 분류하는 방법에는 여러 가지가 있지만, 현재 가장 널리 알려진 방법은 페이로드 시그니처 기반의 분류 방법이다[1,2]. 하지만 페이로드 시그니처 기반의 분류 방법은 기존의 패턴 이외의 새로운 패턴이 나오면 대처하기 힘들다는 한계점을 가지고 있다. 이는 날이 갈수록 응용 트래픽의 패턴이 다양해지고 복잡해지기 때문에 적합하지 않다. 이에 최근에는 인공지능 분야에 널리 이용되는 기계학습을 이용하여 응용 트래픽을 분류하는 방법이 대두되고 있다[1,2]. 이 방법은 새로운 패턴의 응용 트래픽이 나타나도 기존에 학습한 응용 트래픽의 패턴을 바탕으로 기계가 스스로 분류한다. 게다가 기업에서 여러 사용자가 자유롭게 사용할 수 있도록 OpenAI 툴을 제공하기 때문에 쉽게 이용 가능하다는 장점도 있다[1]. 따라서 본 논문에서는 Tensorflow를 이용한 기계 학습 기반의 응용 트래픽 분류 방법을 제안한다. 이 후 제안하는 방법을 바탕으로 설계된 응용 트래픽 분류 모델을 통해 웹 브라우저를 정확하게 분류하는지 확인한다.

본 논문은 서론에 이어 2장 본론에서 Softmax Regreesion을 사용해야 하는 이유와 기계학습 기반의 응용 트래픽 분류 시스템을 간략하게 소개한다. 다음 3 장 실험에서 웹 브라우저(Chrome, Firefox, Internet Explorer, Swing, Whale)의 트래픽을 수집 후, Tensorflow의 Softmax Regression을 이용하여 만든 모델로 이들의 Accuracy를 확인한다. 마지막으로 4장에서 결론 및 향후 연구에 대해 언급한 뒤 본 논문을 마친다.

Ⅱ. 본 론

먼저 본 논문에서 사용할 Softmax Regression은3개 이상의 분류 대상(Multinomial Classification)에 주로 쓰는 분류 기법이다. 분류 대상이 2개일 경우에는 Binary Classification 을 사용하며, Softmax Regression 보다 더 간단한 방법인 Linear Regression으로 분류 가능하다. 하지만 응용 트래픽과 같이 분류할 대상이 여러 가지일 경우 Multinomial Classification을 사용해야 하므로 Softmax Regression을 사용하는 것이 적합하다[1].

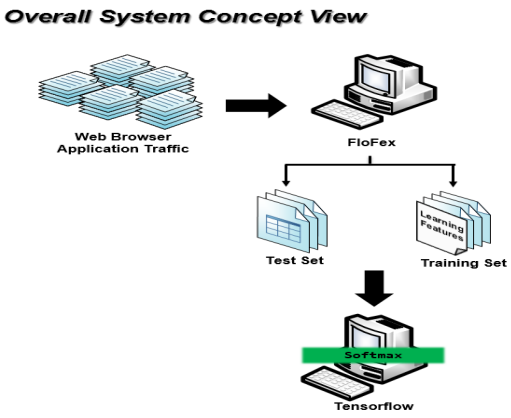

그림 1. 기계학습 기반 응용 트래픽 분류 시스템 구조

이 논문은 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 과학기술인문융합연구사업(No.NRF-2016M3C1B6929228)과 2015년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No.2015R1D1A3A01018057).

본 논문에서 제안하는 기계학습 기반으로 응용 트래픽 분류 시스템 구조는 그림1 과 같다. 먼저 응용 트래픽을 웹 브라우저 마다 수집한다. 이 후 응용 트래픽 학습 특징 추출 프로그램으로 각각의 웹 브라우저 응용 트래픽의 학습 특징을 추출한다. 다음으로 추출한 학습 특징을 바탕으로 Train Set과 Test Set을 만든다. 마지막으로 이전 단계에서 만든 Train Set을 Tensorflow의 Softmax Regression을 이용하여 학습을 한 후 Test Set으로 분류가 잘 되는지 확인한다.

Ⅲ. 실험

위 실험에서 사용한 웹 브라우저는 Chrome, Firefox, Internet Explorer, Swing, Whale이다. 본 실험에서는 5개, 11개 14개의 학습 특징을 사용 하였다. 여기서 Learning Rate는 Train Set을 학습할 때 학습하는 비율을 나타내며, Learning Rate가 그 값에 따라서 분류되는 결과 (Accuracy, Cost) 가 달라진다. Cost는 실제로 분류한 값과 학습을 바탕으로 예측 한 값의 차이를 나타내며 0에 수렴할수록 좋은 결과이다. 반면 Accuracy는 분석된 트래픽 중 올바르게 분류 된 트래픽의 비율을 나타내고 100% 에 수렴할수록 좋은 결과이다. 본 실험에서 사용한 사용할 학습 특징, Learning Rate와 루프 횟수에 따른 결과는 다음과 같다.

표 1. 5 개의 학습 특징을 사용한 분류 결과

LR \ L		2000	20000	200000	10000000
0.01	C	1.545	1.511	1.503	1.478
	A	28.00	30.89	30.44	31.25
0.1	C	1.511	1.503	1.493	1.473
	A	30.89	30.44	29.78	30.28
1.0	C	1.503	1.493	1.479	1.481
	A	30.44	29.78	30.89	30.89
1.5	C	1.502	1.490	1.475	1.491
	A	30.44	29.56	31.56	31.77
3.0	C	1.499	1.486	1.489	1.477
	A	30.44	30.44	32.00	32.03
5.0	C	1.497	1.483	1.464	1.456
	A	29.56	32.00	32.67	33.11

C : Cost / A : Accuracy (%)
LR : Learning Rate / L : 루프 횟수

먼저 표 1은 5개의 학습 특징을 사용하여 분류한 결과이다. 여기서 사용한 학습 특징은 Flow Size, Flow Duration, Packet Count(total forward, backward)이다. 5개의 학습 특징이다. 이를 가지고 Learning Rate와 루프를 도는 횟수를 다르게 하여 실험 한 결과 비교적 낮은 Acciracy가 나타났다. 루프를 2,000번 돌 경우에는 대부분 Accuracy가 30% 대로 나타났다. 그리고 루프를 10,000,000번 돌 경우에도 대부분 Accuracy가 35%대로 크게 차이가 나지 않았다.

다음으로 표 2은 11개의 학습 특징을 사용하여 분류한 결과이다. 여기서 사용한 학습 특징은 Port Number (source, destination), Flow Size, Flow Duration, Packet Count, PPS, FPPS, BPPS (Packet Per Second) 이다. 11개의 학습 특징으로 Learning Rate와 루프를 도는 횟수를 다르게 하여 실험 한 결과 Accuracy는 루프를 도는 횟수에 따라 격차가 크게 나타났다. 루프를 2,000번 돌 경우 대부분의 Accuracy는 50 % 내외지만 10,000,000번 돌 경우에는 대부분 Accuracy가 80 - 85 %로 나타났다.

마지막으로 표 3은 14개의 학습 특징을 사용하여 분류 한 결과이다. 위 실험에서 사용한 학습 특징은 앞의 11개 학습 특징과 Packet with Payload Count (total, forward, backward)을 추가하였다.

표 2. 11 개의 학습 특징을 사용한 분류 결과

LR \ L		2000	20000	200000	10000000
0.01	C	1.337	1.032	0.991	0.851
	A	50.44	49.33	62.89	69.56
0.1	C	1.032	0.909	0.769	0.581
	A	49.33	73.11	77.78	83.23
1.0	C	0.909	0.765	0.583	0.452
	A	72.89	77.78	83.88	84.44
1.5	C	0.961	0.717	0.491	0.447
	A	65.11	81.56	84.00	84.22
3.0	C	0.899	0.635	0.463	0.442
	A	66.89	83.11	83.78	84.44
5.0	C	0.793	0.578	0.448	0.441
	A	56.44	83.78	84.22	84.44

C : Cost / A : Accuracy (%)
LR : Learning Rate / L : 루프 횟수

여기서 Packet Count와 Packet with Payload Count의 차이점은 전자는 각 플로우 안의 전체 패킷들 중 payload가 없는 패킷을 포함한 패킷 수이고, 후자는 각 플로우 안의 전체 패킷들 중 payload가 없는 패킷을 제외한 패킷 수이다. 결과를 보면 루프 횟수를 2,000번 했을 경우에는 대부분 40 - 60 %로 나타났으며 루프 횟수를 10,000,000 했을 경우에는 대부분 90 - 96% 정도의 Accuracy가 나타났다

표 3. 14 개의 학습 특징을 사용한 분류 결과

LR \ L		2000	20000	200000	10000000
0.01	C	1.446	1.139	1.013	0.864
	A	40.22	43.78	62.89	67.13
0.1	C	1.140	1.013	0.733	0.556
	A	43.78	62.89	80.44	85.47
1.0	C	1.013	0.732	0.395	0.249
	A	62.89	80.44	87.78	94.22
1.5	C	0.981	0.670	0.349	0.219
	A	65.11	82.22	89.33	94.67
3.0	C	0.907	0.563	0.262	0.186
	A	67.33	85.33	93.11	95.56
5.0	C	0.837	0.487	0.243	0.172
	A	72.22	85.78	94.22	96.00

C : Cost / A : Accuracy (%)
LR : Learning Rate / L : 루프 횟수

Ⅳ. 결론 및 향후 연구

본 논문은 웹 브라우저의 트래픽을 분류하는 시스템을 제안하였고 실험을 통해 이를 검증하였다. 특히 기계 학습 기반의 Tensorflow를 이용하여 효율적이고 높은 Accuracy로 분류함으로써 그 타당성을 증명하였다.

향후 위 실험에서 사용한 학습 특징보다 더 최적의 학습 특징을 찾아서 더 높은 Accuracy가 나타나도록 실험 할 것이다. 그리고 웹 브라우저 뿐만 아니라 다른 응용 트래픽으로도 실험을 하여 최종적으로 다양한 응용 트래픽을 정확하게 분류 하는 모델을 만들 계획이다.

참 고 문 헌

[1]이성호, 심규석, 구영훈, 김명섭, “ Tensorflow 기계학습 도구를 이용한 응용 트래픽 분류”, Nov, 18, 2016, pp224-225.

[2]박준상, 윤성호, 안현민, 김명섭, “페이로드 시그니쳐 기반 인터넷 트래픽 분류”, May, 15-16, 2014, pp10-14