Hybrid Feature Selection과 Data Balancing을 통한 네트워크 침입 탐지 모델

민병준, 신동규*, 신동일*
세종대학교 컴퓨터 공학과
okminkr@gmail.com, shindk@sejong.ac.kr, dshin@sejong.ac.kr

Network intrusion detection Model through Hybrid Feature Selection and Data Balancing

Byeongjun Min, Dongkyoo Shin*, Dongil Shin*
Dept. of Computer Engineering, Sejong University

요 약

최근 네트워크 환경에 대한 공격이 급속도로 고도화 및 지능화 되고 있기에, 기존의 시그니처 기반 침입탐지 시스템은 한계점이 명확해지고 있다. 이러한 문제를 해결하기 위해서 기계학습 기반의 침입 탐지 시스템에 대한 연구가 활발히 진행되고 있지만 기계학습을 침입 탐지에 이용하기 위해서는 두 가지 문제에 직면한다. 첫 번째는 실시간 탐지를 위한 학습과 연관된 중요 특징들을 선별하는 문제이며 두 번째는 학습에 사용되는 데이터의 불균형 문제로, 기계학습 알고리즘들은 데이터에 의존적이기에 이러한 문제는 치명적이다. 본 논문에서는 위 제시된 문제들을 해결하기 위해서 Hybrid Feature Selection과 Data Balancing을 통한 심층 신경망 기반의 네트워크 침입 탐지 모델을 제안한다. NSL-KDD 데이터 셋을 통해 학습을 진행하였으며, 평가를 위해 Accuracy, Precision, Recall, F1 Score 지표를 사용하였다. 본 논문에서 제안된 모델은 Random Forest 및 기본 심층 신경망 모델과 비교해 F1 Score를 기준으로 7~9%의 성능 향상을 이루었다.

1. 서론

네트워크 침입 탐지 시스템(NIDS)은 네트워크 트래픽을 감시하여 공격 여부를 판단하는 시스템으로, 기존의 NIDS는 시그니처 기반의 탐지 기법이 주를 이루었다. 이는 전문가를 통해 미리 정립된 공격 패턴과의 패턴 매칭을 통해 공격을 탐지한다. 하지만 APT(Advance Persistent Threat) 공격과 같이 위협이 고도화 및 지능화됨에 따라서 트래픽 및 로그에 대한 분석 과정에서의 시간과 비용적 문제가 발생하고 있다. 최근 이러한 문제를 해결하기 위해 기계학습 기반의 탐지 시스템의 연구가 활발하다[1-3].

현실 세계에서 수집되는 많은 데이터들은 클래스 간 균형이 완벽하지 않은 환경이 대부분으로, 특히 침입 탐지 문제에서는 전체 데이터 중 침입 데이터의 비율이 약 1%로 알려져 있다[4]. 이러한 소량의 침입 데이터로 정상학습을 하는 것은 매우 어려우며, 실시간 탐지 위해 사용 가능한 많은 속성 중에서 학습과 관련 있는 특징들을 선별하는 것도 중요하게 다뤄지고 있다.

본 논문에서는 최근 다양한 도메인에서 좋은 결과를 보이고 있는 심층 신경망 모델을 통한 네트워크 침입 탐지 모델을 제안한다. 또한 학습에 중요한 특징들을 선별하기 위하여 Hybrid Feature Selection 기법을 제안하며, SMOTE(Synthetic Minority Over sampling Technique)기법과 RUS(Random Under Sampling) 기법을 통해 데이터 셋의 불균형 문제를 해소하여[5], 소수 클래스들의 탐지율을 개선하였다.

2. 관련 연구

2.1 기계학습 기반 네트워크 침입 탐지 시스템

강승호 외[2]는 NSL-KDD 데이터로부터 Pearson 상관계수 기반의 특징 선택 알고리즘을 제안하였다. 주어진 임계치 이상의 상관계수를 갖는 특징 집합을 그래프 자료구조로 표현한 뒤, 최소 지배 집합(Minimum dominating set)문제로 정의하였다. 최희수 외[3]는 NSL-KDD 데이터로부터 특징들의 빈도수와 평균값을 통한 새로운 특징 선택 기법 AR(Attribute Ratio)을 제안하였다. Nutan 외[4]는

Hybrid Feature Selection 방법을 제안하였다. 서로 다른 특징 선택 알고리즘으로부터 중복제거 합집합으로 표현하여 학습에 사용하였다.

2.2 NSL-KDD 데이터 셋

KDD CUP 99 데이터 셋[6]은 1999년 DARPA 침입탐지 평가 프로그램을 통해 만들어진 데이터 셋으로, 미 공군의 네트워크를 모델링하여 38가지의 네트워크 침입 탐지 공격 시뮬레이션을 통해 만들어졌다. M. Tavallaee 외[6]은 KDD CUP 99 데이터 세트의 규모가 지나치게 크며, 많은 중복 레코드 등을 포함하는 문제점을 지적하며 NSL-KDD 데이터 셋을 제안하였다. NSL-KDD 데이터 세트는 41개의 컬럼으로 구성되며, 표1과 같이 4가지 공격 유형을 포함하고 있다.

<표 1> NSL-KDD 데이터 셋의 공격 유형

공격 유형	설명
DoS	서비스 거부 공격
Probe	침입 전 취약점 분석을 위한 사전 작업
U2R	루트 권한 탈취를 위한 비인가 접근
R2L	원격으로부터의 비인가 접속 시도

3. 본론

3.1 데이터 전처리

NSL-KDD 데이터 셋 컬럼들의 데이터 형식은 nominal, binary, numeric 3가지로 구분할 수 있다. nominal 데이터들은 모두 정수형으로 인코딩 한 뒤 원핫(onehot) 벡터로 변환하였으며, numeric 데이터들에 대해서는 최소 최대 정규화(Min-max Normalization)를 진행하며, binary 데이터들의 경우 모두 0과 1로 구성되기 때문에 별다른 전처리 과정을 수행하지 않는다. 이를 통해 최종적으로 41개의 입력차원이 122개의 입력차원으로 변환되어 학습에 사용된다.

3.2 Hybrid Feature Selection

본 논문에서는 Pearson 기반 특징 선택과 AR 기반 특징선택 및 Feature Importance 기반 특징 선택을 활용한 HFS(Hybrid Feature Selection)을 제안한다. HFS 기법은 그림 1에 명시된 3가지 특징 선택 기법들의 교집합 특징 집합을 사용한다. HFS는 중첩 특징 제거 및 학습에 무관한 특징 제거 2가지 목적에 따라 특징 선택 기법들을 나뉘어 구성하였다. 실제로 Pearson 상관계수가 높은 특징들끼리는 Feature Importance를 뽑을 경우 같이 높은 값을 가지게 된다. 따라서 Feature Importance 만을 통해 특징 선택을 할 경우 이러한 중첩 특징들을 고려할 수 없다. 본 논문에서 제시하는 HFS는 이러한 두 가지 관점의 특징들을 모두 걸러낼 수 있다. 또한 'num_outbound_cmds' 특징은 표준편차가 0으로 관찰되어 사전에 제거하였다.

<표 2> 0.9 이상의 Pearson 상관계수 관계 속성들

	완전 그래프 노드
1	dst_host_srv_count, dst_host_same_srv_rate
2	rerror_rate, srv_rerror_rate, dst_host_rerror_rate, dst_host_srv_rerror_rate
3	serror_rate, srv_serror_rate, dst_host_serror_rate, dst_host_srv_serror_rate
4	num_comproimised, num_root

Pearson 상관계수를 통한 특징선택 기법에서는 NSL-KDD 데이터 셋의 numeric 속성들(31개)의 특징들에 대해서만 진행하였다. 0.9 이상의 상관계수를 가진 특징들 간의 관계를 무방향 그래프 자료구조로 표현한 결과 4쌍의 완전 그래프로 표현되며, 표 2는 4쌍의 그래프들의 노드들을 보여준다. 각 그래프들 사이에서 특징 집합을 대표할 수 있는 최소수의 특징들을 선택하면 특징 벡터의 크기를 최소화 할 수 있다[1]. 하지만 표 2의 결과 그래프는 완전 연결 그래프로 어떠한 것을 임의 삭제하여도 무방하다. 이를 통해 113개의 부분 특징 집합이 선택되었다.

<표 3> HFS를 통해 선택된 특징 집합

특징 집합 (39)
count, diff_srv_rate, dst_bytes, dst_host_count, dst_host_diff_srv_rate, dst_host_same_src_port_rate, dst_host_srv_count, dst_host_srv_diff_host_rate, duration, flag_REJ, flag_RSTR, flag_S0, flag_SF, hot, is_guest_login, logged_in, num_compromised, num_failed_logins, num_file_creations, protocol_type_icmp, protocol_type_tcp, protocol_type_udp, rerror_rate, root_shell, same_srv_rate, serror_rate, service_domain_u, service_eco_i, service_ftp, service_ftp_data, service_http, service_other, service_private, service_smtp, service_telnet, src_bytes, srv_count, srv_diff_host_rate, wrong_fragment

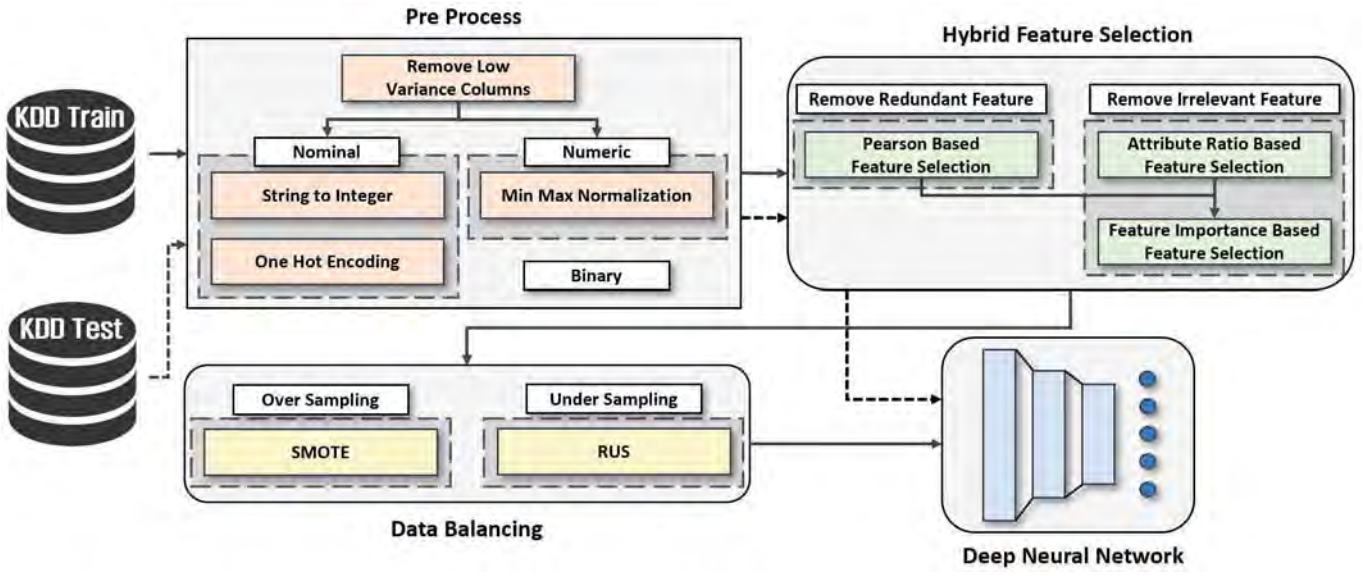

그림 1. 제안하는 네트워크 침입 탐지 모델

AR 기반 특징 선택 기법에서는 0.1 임계값을 이용해 특징들을 선택한 결과 50개의 특징 집합이 선택되었으며, Feature Importance는 Random Forest 모델을 학습시켜 추출하여 상위 55개의 특징 집합을 선출하였다. 해당 특징 선택 기법들의 교집합을 통해 제안되는 HFS 기법의 최종 특징 집합은 표 3과 같이 39개 특징 집합으로, 32% 규모로 축소되었다.

3.3 Data Balancing

불균형 데이터를 통하여 심층 신경망 모델을 학습할 경우는 다수 클래스들에 편향된 학습을 진행하기에 소수 클래스들의 분류 성능은 크게 떨어진다. 본 논문에서는 심각한 불균형 데이터로 분류되는 NSL-KDD 데이터 셋의 불균형 문제를 해소하기 위해서 그림 1에 표기된 SMOTE 기법과 RUS 기법을 활용한다. 표 4를 통해 데이터 불균형을 해소한 데이터 셋의 샘플 수와 비율을 확인할 수 있으며, 다수 클래스는 언더 샘플링을, 소수 클래스는 오버샘플링을 진행하였다.

<표 4> NSL-KDD Dataset 클래스별 샘플

	KDD Train+		Balanced KDD Train+	
Normal	67343	(53%)	45000	(25%)
DoS	45927	(37%)	45927	(25.2%)
Probe	11656	(9.11%)	30000	(16.6%)
U2R	52	(0.04%)	30000	(16.6%)
R2L	995	(0.85%)	30000	(16.6%)
Total	125973		180927	

4. 실험

<표 5> 실험에 사용된 심층 신경망 구조

DNN Parameters	
Architecture	[39-256-512-512-5]
Activation / Initializer	Relu, Softmax / He Uniform
Regularizer / Strength	L2 / 0.0001
Optimizer / Learning rate	Adam / 0.0005
Loss	Cross Entropy

실험에 사용한 심층 신경망의 구조는 표 5와 같으며, 전체 학습 데이터의 20%는 validation set으로 활용하였다. 학습 중 validation loss가 10 epoch 이상 증가할 경우 조기 멈춤을 실행하였으며, 이후 가장 validation loss가 낮은 모델을 선택하여 실험에 사용하였다.

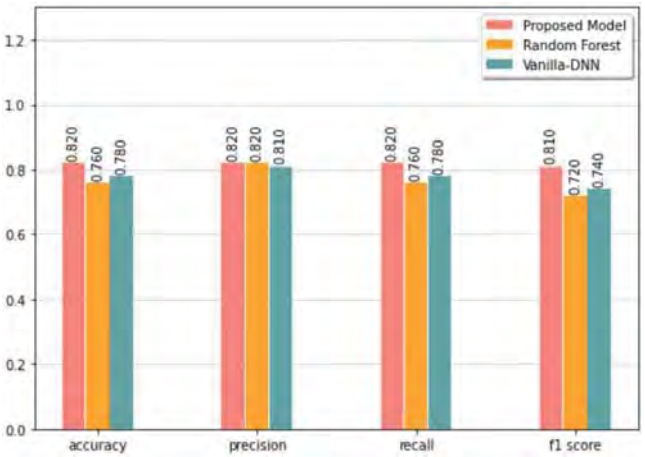

그림 2 제안된 모델의 성능 비교

불균형 데이터에 대해서 Accuracy만을 가지고 평가하는 것은 부적합하다. 따라서 본 논문에서는 학습된 모델의 성능 평가로 Accuracy, Precision, Recall, F1 Score를 평가 지표로 사용하였다. 그림 2를 참조하면 제안된 모델의 4가지 성능지표가 비교 모델들에 비해 더 좋은 것으로 확인된다.

<표 6> 제안된 모델의 클래스별 성능 평가

Proposed Models				
	precision	recall	f1	support
DoS	0.96	0.85	0.90	7458
Probe	0.84	0.66	0.74	2421
R2L	0.64	0.42	0.51	2754
U2R	0.26	0.14	0.18	200
Normal	0.77	0.96	0.86	9711
Total	0.82	0.82	0.81	22544

<표 7> 비교 모델들의 소수 클래스 분류 결과

Random Forest				
	precision	recall	f1	support
R2L	0.99	0.08	0.15	2754
U2R	0.5	0.01	0.02	200
Vanilla DNN				
R2L	0.95	0.08	0.14	2754
U2R	0.	0.	0.	200

표 6은 본 논문에서 제안된 모델의 각 클래스별 성능 지표로, 표 7과 비교하면 소수 클래스들의 성능 지표들을 비교할 수 있다. 특징 선택과 데이터 균형을 맞추지 않은 데이터를 학습한 두 모델은 소수 클래스들에 대해서 제대로 분류를 하지 못하고 있는 것을 확인할 수 있다. 단순 심층 신경망의 모델의 경우 네트워크 구조는 동일하지만 모든 U2R 클래스의 분류에 실패한 것을 확인하였다. 그에 반해 본 논문에서 제안된 모델은 R2L 클래스의 탐지 성능의 경우 눈에 띄게 상승한 것이 보이며, U2R 클래스 탐지 성능 또한 개선이 된 것을 확인할 수 있다. 그럼에도 불구하고 U2R의 분류 성능이 낮은 것을 확인할 수 있는데, 이는 Train 셋에서 제공되는 데이터양이 52개인 반면에 Test 셋에서 제공되는 양은 200개로 더 적은 것을 알 수 있다. 따라서 테스트 데이터들의 분포가 더 넓게 분포 되어 있을 것으로 분석된다.

5. 결론

본 논문에서는 네트워크 침입 탐지의 성능 개선을 위해 Hybrid Feature Selection 기법을 제안하였다. 또한 학습에 사용된 NSL-KDD 데이터 셋의 불균형 문제를 해소하여 성능이 개선된 네트워크 침입 탐지 모델을 제안하였다. 특징 추출을 통해 32% 규모로 입력 차원을 축소할 수 있었으며, 오버 샘플링 기법을 통해 소수 클래스들의 성능 개선을 실험을 통해 확인할 수 있었다. 이를 통해 본 논문에서 제안한 모델의 F1 Score 가 두 모델에 비해서 7~9% 높은 것으로 확인할 수 있었다. 향후 연구로는 VAE, GAN과 같은 생성모델들을 통해 데이터 증감(Data Augmentation)기법을 연구하여, SMOTE를 대체할 수 있으며. 또한 다른 침입 탐지 데이터 셋을 통한 연구 또한 진행할 수 있다.

Acknowledgement

본 연구는 방위사업청과 국방과학연구소의 지원으로 수행되었습니다(UD190016ED).

참고문헌

[1] 강승호, 정인선, and 임형석. "실시간 공격 탐지를 위한 Pearson 상관계수 기반 특징 집합 선택 방법." 융합보안논문지 18.5 (2018): 59-66.

[2] Chae, Hee-su, et al. "Feature selection for intrusion detection using NSL-KDD." Recent advances in computer science (2013): 184-187.

[3] Haq, Nutan Farah, Abdur Rahman Onik, and Faisal Muhammad Shah. "An ensemble framework of anomaly detection using hybridized feature selection approach (HFSA)." 2015 SAI Intelligent Systems Conference (IntelliSys). IEEE, 2015.

[4] Song, Jungsuk, et al. "Correlation analysis between honeypot data and IDS alerts using one-class SVM." Intrusion Detection Systems (2011): 173-192.

[5] Yang, Xin-Li, et al. "High-impact bug report identification with imbalanced learning strategies." Journal of Computer Science and Technology 32.1 (2017): 181-198.

[6] Tavallaee, Mahbod, et al. "A detailed analysis of the KDD CUP 99 data set." 2009 IEEE symposium on computational intelligence for security and defense applications. IEEE, 2009.