기계학습 기반 네트워크 정상행위 모델링에 관한 연구

권성문*, 손태식**
*아주대학교 컴퓨터공학과
**아주대학교 사이버보안학과
e-mail:calmcombat@gmail.com
e-mail:tsshon@ajou.ac.kr

A Study on Modeling Network Normal Behavior based on Machine Learning

Sungmoon Kwon*, Taeshik Shon**
*Dept of Computer Engineering, Ajou University
**Dept of Cyber Security, Ajou University

요 약

네트워크 정상행위 모델링이란 대상 네트워크 및 시스템에서 동작 가능한 행위 중 허용된 행위를 모델링하는 것을 의미한다. 정상행위 모델은 해당 모델의 정상 이외 범주의 알려지지 않은 비정상 행위의 탐지 가능성을 가지고 있어 활용도가 높다. 네트워크 및 시스템의 복잡도가 증가할수록 특성의 파악이 힘들며 이로 인해 주요 특징의 누락이 발생할 수 있어 대상 네트워크의 다수의 데이터에 기반한 기계학습 기반의 네트워크 정상행위 모델링에 관한 다양한 연구가 진행되고 있다. 본 논문에서는 딥러닝을 포함하여 네트워크 정상행위 모델링에 사용될 수 있는 다양한 기계학습 기반의 기법을 제시한다.

1. 서론

네트워크 정상행위 모델링 또는 단일 분류 모델이란 대상 네트워크 및 시스템에서 허용된 행위를 모델링하는 것을 의미한다. 이와 반대로 악성 행위의 시그니쳐에 기반한 비정상행위 모델은 모델 학습에 사용한 악성 행위 또는 유사 악성행위만을 탐지 가능한 한계점이 있다. 그러나 정상행위 모델의 경우 정상행위 이외의 행위는 모두 악성으로 규정짓기 때문에 아직 알려지지 않은 비정상행위의 탐지의 가능성을 내포하고 있다. 반면 정상행위 모델의 한계점은 모델이 정상행위 만의 고유 특징을 포함하지 못할 경우 비정상행위 또한 정상으로 분류될 수 있으며 반대로 모델이 특정 정상행위의 특징만을 포함한 경우 정상행위가 비정상으로 분류될 수 있는 점이 있다. 네트워크 정상행위 모델링을 위해서는 네트워크 및 시스템이 단순한 경우 화이트리스트 규칙 또는 상태 다이어그램을 작성하는 것으로 모델링이 가능할 수 있다. 그러나 대상의 복잡도가 증가할수록 특성 파악이 힘들며 이로 인해 파악되지 못한 주요 특징의 누락이 발생할 수 있다. 따라서 정상행위로 가정되는 많은 네트워크 데이터를 사용하여 기계학습 기반의 네트워크 정상행위 모델링이 연구되고 있으며 최근에는 딥러닝 또한 활용되고 있다. 본 논문에서는 딥러닝을 포함하여 네트워크 정상행위 모델링을 하기 위한 다양한 기계학습 기반의 기법을 제시한다.

본 논문의 2장에서는 정상행위 모델링의 문제점을 설명하며 3장에서는 정상행위 모델링 기법들을 설명한다. 마지막 4장에서는 결론 및 향후 연구로 논문을 맺는다.

2. 정상행위 모델링의 문제점

정상행위 모델링은 정상행위만으로 학습을 수행하여 일반적으로 분류모델에 사용되는 알고리즘을 바로 사용하지 못한다. 군집화(clustering)를 통해 정상 데이터를 각 특징을 가진 클래스로 나누어도 비정상행위 인풋을 하나의 정상 클래스로 분류할 뿐이다. 따라서 정상행위 모델링을 위해서는 추가적인 기법을 사용하거나 변형 알고리즘을 사용해야 한다. 분류 모델로 높은 정확도를 가진 SVM(support vector machine)의 변형인 OCSVM(one class support vector machine)은 다양한 분야에서 적용 가능함을 보였다.[1][2] 그러나 OCSVM이 SVM에 비해 정확도가 낮은 것은 사실이며 네트워크 분야에서는 단 1%의 오탐률도 수십만 패킷에서 수천 건의 거짓 경보를 뜻하기 때문에 정상적인 활용이 힘들다.

반면 정상행위 데이터에 의존하기 때문에 정상행위 데이터에 비정상데이터가 포함되어 있을시 정상적인 모델링이 힘들다. 이를 위해 아웃라이어(outlier)를 식별하고 이에 대한 제거가 가능한 경우 보다 정확성이 높은 정상행위 모델링이 가능하다.

3장에서는 이러한 문제점을 고려하여 네트워크 정상행위 모델링에 활용 가능성이 있는 기법들을 제시한다.

3. 정상행위 모델링 기법

3.1 K-means, State diagram, Apriori

K-means 알고리즘은 대상 네트워크 또는 시스템의 특성 정보가 없는 경우 유사 데이터를 군집화 함으로써

대상의 특성 정보를 추출할 수 있다. 또한 유사 데이터 군에 벗어난 아웃라이어를 식별 가능한 장점도 있다. 그러나 K-means 모델 자체는 데이터의 정상/비정상 판단은 수행할 수 없다. 단, 입력 값에 대한 K-means 모델 출력 값의 흐름을 상태 다이어그램으로 나타내어 [3]과 같은 dynamic bayesian network(DBN)이나 마코프 체인방법론을 이용하여 분류 모델로 활용할 수 있다.

Apriori 알고리즘은 각 데이터의 속성(attribute) 값들의 상관관계를 분석하는 알고리즘이다. 제어시스템과 같이 규칙적인 시스템에서 Apriori 알고리즘을 사용할 시 1에 가까운 신뢰도를 가지는 규칙만을 사용하여 정상행위 모델링을 수행할 수 있다. 또한 신뢰도가 낮은 규칙을 분석하여 아웃라이어를 식별할 수 있다. 그러나 규칙성이 보장되지 않는 환경에서는 신뢰도가 높은 규칙을 기대하기 어려우며 데이터의 상관관계 분석에 활용 가능할 것이다.

K-means와 Apriori의 경우 정상행위 모델링을 위한 방법 외에도 대상 네트워크 및 시스템의 특성을 파악하기 위한 사전 단계로도 활용 될 수 있을 것이다.

3.2 딥러닝 기반 정상행위 모델링 기법

현재 사용되는 딥러닝 기술의 대부분은 학습을 수행한 다음 마지막 계층으로 분류 계층을 붙여 분류 모델로 사용하게 된다. 따라서 정상행위만으로 학습하는 경우 인공신경망의 은닉 계층의 가중치(weight)와 무관하게 분류 계층에서 정상 클래스의 가중치가 증가하는 방식으로만 학습되기 때문에 활용 가능한 모델의 생성을 기대하기 어렵다. 따라서 본 논문에서는 CNN(convolution neural network)과 같은 분류 모델로써 높은 정확도를 가지는 모델 외에 RNN(recurrent neural network)과 AE(autoencoder)를 활용하여 정상행위 모델링을 하는 기법에 대해 제시한다.

3.1.1 RNN 기반 기법

RNN은 현재 데이터를 예측하기 위해 이전 데이터를 사용한다. RNN의 가장 큰 장점은 다음의 데이터를 예측하는데 있다. 따라서 직전의 데이터를 통해 생성한 예측 데이터와 현재 취득한 데이터의 차이 값을 통해 현재 취득한 데이터가 정상인지 비정상인지 판별하는 모델을 생성하는 것이 가능하다. 정상과 비정상을 판별하는 차이 값은 특정 임계값을 사용하거나 베이지안(bayesian)과 같은 확률 모델을 통해 정확도를 높일 수 있다. 또한 데이터의 주기성을 미리 파악하여 입력 값으로 활용하거나 패턴성이 없는 독립적인 속성 데이터를 제외할 경우 정확도 향상을 기대할 수 있다.

3.1.2 AE 기반 기법

AE는 입력 데이터와 유사한 출력 데이터를 생성하는 모델로 단순히 보자면 다변수 선형 회귀 분석(multivariable linear regression)과 유사하다. AE는 학습 단계에서 각 입력 데이터에 대한 출력 데이터간의 차이 값인 재구축 에러(reconstruction error)의 평균이 최솟값을 가지도록 학습이 수행되며 그 결과 재구축 에러 값은 학습 방법에 따라 상이할 수 있지만 표준 정규 분포와 같은 고른 분포를 가지게 된다.[4] 따라서 학습이 완료된 AE 모델은 특정 입력 데이터에 대한 재구축 에러의 Z score에 근거하여 해당 데이터의 정상/비정상을 판단할 수 있는 확률적 모델로 활용할 수 있다.

4. 결론 및 향후 연구

본 논문에서는 네트워크 정상행위 모델링의 문제점을 설명하고 이에 대응하기위한 정상행위 모델링 맞춤형 기법들을 제시하였다. 제시한 기법에는 고전적인 K-means, Apriori 알고리즘과 상태 다이어그램에 기반한 모델링 기법과 RNN, AE의 특성을 이용한 기법을 제시하였다. 사전에 파악한 특성 정보가 없는 경우 K-means 알고리즘이 모델링의 시작으로 유용하며 데이터의 분포가 밀집된 경우 AE 기반 기법이 우수한 성능을 보일 수 있다. 제어시스템과 같이 정상행위가 규칙적으로 분명한 경우 Apriori 알고리즘이 적합한 선택일 수 있으며 네트워크 통신 패턴의 주기 특성을 파악하고 있을시 RNN이 적합한 선택이 될 수 있을 것이다. 이와 같이 각 알고리즘을 통해 효과적인 모델링을 하기 위한 전제조건이 다르며 각 모델의 네트워크 정상행위 모델링에 유효성을 검증하는 것을 향후 연구로 수행할 예정이다.

ACKNOWLEDGEMENT

"본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음" (IITP-2018-2016-0-00304)

참고문헌

[1] Y. Chen, X. S. Zhou, T. Huang, One-class svm for learning in image retrieval, in: International Conference on Image Processing, Thessaloniki, Greece, 2001.

[2] H. Alashwal, S. Deris, R. Othman, One-class support vector machines for protein-protein interactions prediction, International Journal of Biomedical Sciences 1 (2) (2006) 120-127.

[3] Yoon, Man-Ki, and Gabriela F. Ciocarlie. "Communication pattern monitoring: Improving the utility of anomaly detection for industrial control systems." NDSS Workshop on Security of Emerging Networking Technologies. 2014.

[4] Aygun, R. Can, and A. Gokhan Yavuz. "Network anomaly detection with stochastically improved autoencoder based models." Cyber Security and Cloud Computing (CSCloud), 2017 IEEE 4th International Conference on. IEEE, 2017.