GCN을 적용한 TTP 사이버 데이터 기반 국가 예측 방안
The proposed approach for country prediction with TTP-based cyber data using GCN

신성욱 · 신찬호 · 서성연 · 이인섭 · 최창희
Sunguk Shin · Chanho Shin · Seongyun Seo · Inseop Lee · Changhee Choi

국방과학연구소
(ssw1419@add.re.kr)

ABSTRACT

Cyber attacks are organized and constantly increasing. In this reason, it is important to analyze cyber attacks. Cyber kill chain is used to remove threats by analyzing cyber threats in each step. The cyber kill chain can be constructed in a graph because technologies are connected. There are various ways to analyze the graph. Graph Convolution Network is a deep learning technique for analyzing graphs, and it can be used to analyze cyber kill chain and propose to predict attack groups.

Key Words : Deep Learning, GCN, TF-IDF, TTP, ATT&CK

1. 서론

사이버 위협은 점점 다양해지고 발전하면서 국가 단위의 공격도 발전하고 있다. 사이버 공격이 선형적으로 진행된다는 점을 활용하여 탐지하고 대응하기 위해 사이버 킬체인이 등장하게 되었다. MITRE[1]는 사이버 킬체인을 계층 구조로 분류하여 위협을 분석할 수 있는 프레임워크를 제공하고 있다.

현재는 이러한 위협에 대해 룰베이스를 기반으로 실시간 탐지를 하며 차단 및 분리 등의 대응을 수행하고 있다. 하지만 딥러닝이 발전함에 따라 예측하는 문제에 대한 정확도가 증가하고 있기 때문에 룰베이스 기반으로 실시간으로 탐지를 한 뒤 다양한 정보를 예측할 수 있게 되었다.

선형적인 구조를 분석하고 예측하는 방법으로 RNN이나 LSTM과 같은 딥러닝 기법이 등장하게 되었다. 킬체인은 위협이 시간적인 순서로 이루어져 있기 때문에 선형적인 구조를 띄고 있지만 연관 관계를 살펴보면 그래프의 형태를 가지게 있기 때문에 단순 선형 구조로 사이버 위협을 분석하는 것보다 그래프의 특성을 유지하여 분석을 할 필요가 있다.

본 논문에서는 그래프 구조를 분석하는 GCN[2]과 그래프를 구성하는 노드들의 벡터를 구할 수 있는 TF-IDF를 기술하고 해당 기법을 적용하였을 때 분류 문제를 실험함으로써 사이버 영역에서의 활용 가능성을 제시한다.

2. background

2.1 Term Frequency Inverse Document Frequency (TF-IDF)

TF-IDF는 특정 문서에서 특정 단어의 등장 횟수인 TF 값과 특정 단어 t가 등장한 문서의 수의 역을 의미하는 IDF 값의 곱을 의미한다. TF는 문서에 대한 벡터를 의미하며 IDF는 단어데 대한 스칼라 값을 가지고 있기 때문에 TF-IDF는 문서에서 단어에 대한 중요도를 포함한 벡터값이 된다.

2.2 Graph Convolution Network (GCN)

GCN은 그래프 정보를 활용하여 convolution 연산을 수행하는 딥러닝 기법이다. 그래프를 이미지처럼 연산할 수 있도록 인접행렬과 피처행렬을 활용한다. 인접행렬은 node와 edge로 정보를 가지고 있는 행렬을 의미하고 피처행렬은 각 노드가 가진 고유한 벡터를 의미한다. 인접행렬과 피처행렬이 GCN에 입력으로 들어가 인접행렬×피처행렬×웨이트를 계산한다.

2.3 MITRE TTP

TTP는 Tactic, Technique, Procedure의 약자로 MITRE에서 사이버 공격을 분류하였다. Tactic은 사이버 공격의 목적을 의미하고 Technique은 tactic을 수행하기 위해 사용할 수 있는 단위의 기법을 의미한다. 실험을 위해서 MITRE 버전은 5.2를 사용하였다.

3. 데이터셋 및 모델

사용한 데이터셋은 rcATT[3]의 training data를 활용하였다. 해당 데이터는 보고서와 보고서에 해당하는 tactic과 technique 정보를 포함하고 있다. 데이터셋에서 제공하는 tactic과 technique 정보는 단순히 0과 1로 표현되어있기 때문에 그래프로 구성하기 위해 다음과 같이 설정하였다. Tactic 노드는 MITRE에서 제공한 tactic의 순서로 나열한 뒤 연결하였다. 같은 tactic내의

technique들은 완전그래프로 구성하였으며 인접한 tactic 내의 technique 들도 모두 연결하였다.

데이터셋의 보고서를 통해 예상되는 위협 그룹 1과 그룹 2인 경우 65개를 사용하였으며 학습 데이터는 그룹 1 24개와 그룹 2 22개, 테스트 데이터는 그룹 1 10개와 그룹 2 9개를 활용하였다.

Technique과 tactic 노드의 feature 정보를 추출하기 위해서 TF-IDF를 활용하였다. MITRE는 tactic과 technique에 대한 설명을 제공하고 있기 때문에 해당 설명을 문서로 취급하여 벡터화시켰다.

그래프를 통해 분류하는 문제는 노드 및 엣지 간의 연관관계를 딥러닝으로 학습하여 예측하는 것이기 때문에 모든 데이터 내에 2번 이하로 등장한 노드를 제거하였으며 연결이 자기 자신 한 개만 존재하는 그래프도 제거하고 학습을 진행하였다. 해당 노드를 제거하고 TF-IDF를 통해 생성된 TTP에 대한 벡터는 1973의 크기를 가진다.

모델은 graph convolution 계층 2개와 linear 계층 1개로 구성되어있다. 데이터가 많지 않기 때문에 과적합을 방지하기 위해서 graph convolution 계층 사이에 dropout을 추가하였다.

4. 실험 결과

학습셋을 통해 20 epoch 학습하였을 때 학습셋에 대한 정확도가 93.48%일 때 테스트셋에 대한 정확도가 73.68% 나왔다. 예측한 결과에 대한 confusion matrix는 표 1에 나와 있다. 그룹 1과 그룹 2를 예측하는 문제에서 19 데이터 중 14개를 올바르게 예측하였다. 딥러닝의 특성 상 데이터 학습 입력순서 및 초기 웨이트에 따라 모델의 정확도가 달라지기 때문에 5번의 실험을 진행해보았을 때 비슷한 결과를 얻을 수 있었다.

Table 1. 예측 confusion matrix

	predict 그룹 1	predict 그룹 2
actual 그룹 1	8	2
actual 그룹 2	3	6

표 1에 대한 하이퍼 파라메터로 graph convolution의 출력은 각각 32와 8이며 Adam의 학습률을 0.01로 설정하였다. graph convolution 출력 사이즈와 Adam의 학습률을 변화시켰을 때 실험한 결과 84.21%까지 증가하였으나 반복적인 실험에서 불안정하였기 때문에 표 1에 해당하는 파라메터가 가장 적절한 것을 확인할 수 있었다. 그림 1의 파란색 선은 ROC curve를 의미한다. AUC는 0.7444이다.

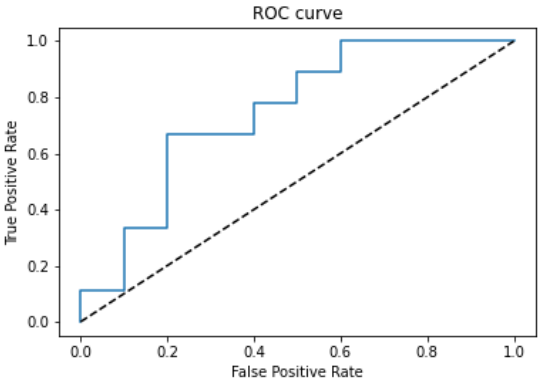

Fig. 1. ROC curve

테스트셋에 대한 정확도가 크게 높지 않은 이유로 2가지가 있다. 첫 번째는 학습 시 데이터의 개수가 46개로 딥러닝 모델을 학습하기엔 매우 적은 숫자이다. 두 번째는 TF-IDF를 통해 생성된 벡터(피처행렬)과 인접행렬이 희소행렬이기 때문에 학습하는데 어려움이 있다.

5. 결론

본 논문에서는 사이버 위협을 그래프로 분석하는 방법에 대해서 서술하였다. 기존의 위협을 탐지하는 것을 넘어서 위협에 대한 그래프 정보를 가지고 그룹를 예측할 수 있다는 것을 보였으며 적은 데이터를 가지고도 그래프를 통해 예측이 가능하다는 것을 확인할 수 있었다.

노드의 피처를 추출하는 방법에 TF-IDF를 사용하였지만 수동으로 노드간의 연관관계를 분석하여 피처행렬을 구성하여 더 좋은 성능을 기대할 수 있을 것이며 TTP 데이터를 활용하여 그룹을 예측하는 것뿐만 아니라 다음 행동 혹은 목적 등을 예측하는 실험을 진행할 수 있을 것이다.

References

[1] The MITRE Corporation, attack.mitre.org, 2015-2021.

[2] Kipf, Thomas N., and Max Welling. “Semi-supervised classification with graph convolution networks.” International Conference on Learning Representations (ICLR), 2017.

[3] Legoy, Valentine, et al. “Automated Retrieval of ATT&CK Tactics and Techniques for Cyber Threat Reports.” arXiv preprint, arXiv:2004.14322, 2020.