그래프 기반 웹사이트 간 비밀 번호 유사도 예측 모델 연구

이승환, 이준석, 최형기*
성균관대학교

pea040119@g.skku.edu, sam1231@skku.edu, meosery@skku.edu

Graph-Based Prediction Model for Password Similarity Across Websites

Seunghwan Lee, Junseok Lee, Hyoung-Kee Choi*
Sungkyunkwan Univ.

요 약

전 세계적으로 비밀번호 유출이 빈번해지면서 credential stuffing 공격 위험이 심각해지고 있다. 본 논문에서는 비밀번호 유사도 그래프를 활용해 서로 다른 웹사이트 간 사용자의 유사 비밀번호 사용 가능성까지 포괄적으로 예측하는 새로운 그래프 기반 모델을 제안한다.

Ⅰ. 서론

전 세계적으로 대규모 비밀번호 유출로 탈취된 비밀번호가 타사이트에서 재사용되어 심각한 보안 위협이 대두되고 있다. 이를 해결하기 위해, 노출된 비밀번호 보유 여부를 확인하는 credential compromise checker 서비스와 사전 예방적 재사용 위험 예측 기법이 제안·연구되고 있다.

그중 PassREfinder 는 그래프 기반 위험 예측 모델로, 웹사이트를 노드로, 비밀 번호 재사용 여부를 간선으로 정의하여 사이트 간 비밀 번호 재사용 가능성을 그래프로 표현한다 [1]. 이후 Graph Neural Network (GNN)를 이용한 간선 예측 기법을 통해 credential stuffing 공격의 위험도를 추정한다.

본 논문에서는 비밀 번호의 단순 재사용 예측을 넘어, 서로 다른 웹사이트 간 사용자가 유사한 비밀 번호를 사용할 가능성까지 포괄적으로 예측할 수 있는 새로운 모델을 제안한다. 제안 모델의 유효성은 다양한 실험을 통해 검증하였으며, 실험 결과를 바탕으로 향후 연구를 위한 개선 방향을 제시한다.

Ⅱ. 본론

1. 그래프 구조

비밀 번호 유사도 그래프는 웹사이트 데이터셋에서 추출한 사이트 간 비밀 번호 유사도 정보를 기반으로 구성된다. 문자열 유사도 계산 알고리즘으로 Jaro-distance [2]를 채택하여, 매칭된 문자와 전위된 문자 수를 활용해 0 과 1 사이의 유사도 값을 산출한다. Jaro-distance 는 매칭된 문자 정보를 중심으로 계산되므로, 비밀 번호 유사도 계산에 적합하다. 그래프의 간선은 동일 사용자에 속하는 두 사이트 간 평균 유사도가 설정된 임계값을 초과할 때에만 형성된다. 또한, 그림 1 과 같이 각 노드는 URL, 카테고리, 국가, IP 주소 및 비밀 번호 구성의 최소 요건을 나타내는 보안 수준으로 구성된 특성 벡터로 정의하였다.

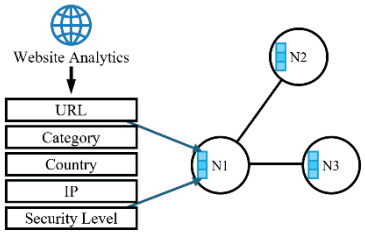

그림 1: 그래프 구조

모델은 transductive 와 inductive setting 학습 환경을 제공한다. Transductive setting 은 모든 웹사이트를 이용하여 그래프를 구성한 뒤, 간선을 분류하여 기존에 존재하는 웹사이트 간의 유사한 비밀 번호를 사용하는지 확인할 수 있다. Inductive setting 은 웹사이트를 분류한 뒤 subgraph 를 구성함으로써, 신규 웹사이트에서도 유사한 비밀 번호를 사용하는지 확인할 수 있다.

2. 모델 설명

본 논문에서는 비밀 번호 유사도 그래프를 학습하기 위해 네 가지의 모델을 사용한다.

MLP (Multi-Layer Perceptron). 그래프 토폴로지를 고려하지 않고 각 노드의 특성만을 입력으로 받아 학습하는 완전 연결 신경망이다[3]. 그래프 구조 정보를 활용하지 않으므로, 웹사이트의 기본 속성 학습을 위한 단순 벤치마크 모델로 사용된다.

GCN (Graph Convolutional Network). 합성곱 연산을 통해 각 노드가 이웃 노드의 정보와 자체 특성을 동시에 학습하는 모델이다[4]. 본 연구에서는 비밀 번호 유사도 그래프에서 이웃 노드 특성을 평균 또는 정규화된 합으로 집계한 후 이를 기반으로 노드 임베딩을 학습한다.

GAT (Graph Attention Network). 어텐션 메커니즘을 도입하여 각 이웃 간 가중치를 학습함으로써, 중요한 이웃으로부터 전파되는 정보를 선택적으로 강화하는 모델이다 [5]. 이를 통해 비밀 번호 유사도 그래프에서 중요한 이웃의 영향력을 선택적으로 학습하는 GCN 변형 모델이다.

GraphSAGE. 대규모 그래프에 대한 미니배치 학습을 지원하며, 이웃 집계 함수 (Aggregator)를 학습 가능한 형태로 일반화한 모델이다 [6]. 이 모델은 aggregator 를 통해 이웃 노드의 특성을 집계함으로써 특성 표현력을 강화한다. Aggregator 로는 세가지의 함수를 적용한다. mean aggregator 는 이웃 노드 임베딩의 단순 평균을, attention aggregator 는 학습된 attention 가중치를 이용해 이웃 간 중요도 차이를 반영한 가중 평균을 수행한다. no hidden aggregator 는 선형 변환이나 비선형 활성화 없이 이웃 임베딩을 합산 또는 평균하여 집계한다.

3. 실험 방법

Dataset. 자격증 판매 서비스 유출 사고에서 수집된 데이터셋인 Cit0day [7]를 사용한다. 총 22,378 개의 웹사이트 중에서 접속 가능한 1,653 개의 웹사이트 데이터를 확보하여 1,653 개의 웹사이트 노드와 253,609 개의 edge 로 비밀 번호 유사도 그래프를 구성한다.

모델에 따른 성능 평가. 단순 신경망인 MLP 모델과, 이웃 노드의 특징을 활용하는 GCN 모델을 학습하였다. 아울러 GCN 의 구조적 한계를 보완한 GAT 및 GraphSAGE 모델을 동일한 조건 하에서 추가로 학습함으로써 네 가지 모델 간 성능 차이를 비교한다.

Aggregator 에 따른 성능 평가. GraphSAGE 모델에서는 이웃 노드 집계 방식에 따라 학습 결과가 달라지므로, mean, attention, no hidden aggregator 를 적용하여 실험을 수행하였다. 각 방식이 모델의 성능에 미치는 영향을 분석하여 최적의 집계 방법을 도출한다.

Ⅲ. 실험 결과 및 분석

1. Model 에 따른 성능

표 1: MLP, GCN, GAT, GraphSAGE 모델의 성능 비교

Model	Transductive		Inductive	
	Accuracy	F1-score	Accuracy	F1-score
MLP	0.8797	0.8385	0.7369	0.6521
GCN	0.8124	0.7487	0.7445	0.4546
GAT	0.8714	0.8227	0.7456	0.6147
GraphSAGE	0.8020	0.6951	0.7336	0.5938

표 1 에서 확인할 수 있듯이 본 실험에서는 네 모델의 분류 성능을 평가하였다. City0day 전체 데이터에 비하여 학습에 사용된 데이터는 제한적이므로, 단순 구조의 MLP 모델이 높은 성능을 보인다. GCN 모델은 spectral filterin 에서 노드 간 표현이 지나치게 균일해져 성능이 저하되었다. GAT 모델은 edge 별 가중치로 중요한 이웃만을 선별하여 과도한 스무딩을 완화하여 높은 성능을 보였다. GraphSAGE 모델에서는 이웃 집계 과정에서 핵심 특징이 희석되어 성능이 저하되었다.

소규모 그래프에서는 단순 구조 모델이 과도한 복잡성 없이 안정적으로 작동하며, 새로운 노드 예측에서는 attention 메커니즘이 정확도를 향상시킨다. GAT 모델의 self-attention 구조를 다중 헤드 어텐션이나 동적 가중치 보정 기법으로 확장하여 성능 개선 방안을 제시한다.

2. Aggregator 에 따른 성능

표 2: GraphSAGE 모델의 aggregator 의 성능 비교

Aggregator	Transductive		Inductive	
	Accuracy	F1-score	Accuracy	F1-score
Mean	0.8289	0.7623	0.7639	0.6651
Attention	0.8020	0.6951	0.7336	0.4546
No hidden	0.8226	0.7600	0.7100	0.6327

표 2 에서 확인한 바와 같이 본 실험에서는 GraphSAGE 모델의 aggregator 함수를 변경하여 성능을 평가하였다. 데이터셋으로 사용된 웹사이트가 적기 때문에 단순한 구조를 가진 mean 과 no hidden 에 비하여 복잡한 구조를 가진 attention 의 경우 다른 aggregator 에 비하여 성능이 저하되는 모습을 보인다.

제한된 학습 데이터 환경에서 파라미터가 적은 mean 과 no hidden aggregator 가 안정적인 예측 성능을 보였다. 모델 성능 향상을 위해서 대규모 그래프 적용 및 다른 집계 함수를 사용하여 모델의 성능을 개선이 가능하다.

Ⅳ. 결론

본 논문에서는 웹사이트 간에 사용자가 유사한 비밀번호를 사용할 가능성을 그래프 기반의 위험 예측 모델을 제안했다. 총 네가지 모델 MLP, GCN, GAT, GraphSAGE 를 실험하고 분석했다. 또한 GraphSAGE 의 mean, attention, no hidden aggregator 의 성능을 비교하였다. 향후 대규모 유출 데이터 활용 및 높은 성능을 보인 GAT 모델의 self-attention 구조를 확장하여 성능 개선을 진행할 계획이다.

ACKNOWLEDGMENT

이 논문은 2024 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구 결과 임 (No. RS-2024-00398745, 디지털 환경에서의 증거인멸행위 증명 및 대응 기술 개발).

참 고 문 헌

[1] Kim J.; Song M.; Seo M.; Jin Y.; Shin S. “PassREfinder: Credential Stuffing Risk Prediction by Representing Password Reuse between Websites on a Graph,” in Proceedings of the 45th IEEE Symposium on Security and Privacy, pp. 1385-1404, 2024.

[6] M. A. Jaro, “Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida,” *Journal of the American Statistical Association*, vol. 84, no. 406, pp. 414-420, Jun. 1989.

[3] Rumelhart D. E.; Hinton G. E.; Williams R. J. "Learning representations by back-propagating errors," Nature, pp. 533-536.

[4] Kipf T. N.; Welling M. "Semi-Supervised Classification with Graph Convolutional Networks," International Conference on Learning Representations.

[5] Veličković P.; Cucurull G.; Casanova A.; Romero A.; Liò P.; Bengio Y. "Graph Attention Networks," International Conference on Learning Representations.

[6] Hamilton W. L.; Ying Z.; Leskovec J. "Inductive Representation Learning on Large Graphs," Advances in Neural Information Processing Systems, pp. 1024-1034.

[7] “Cit0Day Collection,” Raidforums, https://raidforums.com/Thread-Cit0Day-Collection-Leaked-Download, accessed: 2021-3-13.