[웹 방화벽 로그 분석을 통한 공격 분류] 정리

요약

웹 방화벽(WAF) 로그 데이터를 활용하여 머신러닝 알고리즘으로 공격 유형 식별

`TF-IDF`: 전통적인 텍스트 마이닝 기법으로 단어의 중요도를 평가하는 기술

`BERT`: 문맥을 고려한 단어의 의미 파악에 초점을 맞춘 기술

`AutoML`: 데이터 전처리, 모델 선택, 하이퍼파라미터 조정 등의 과정과 최적화를 자동으로 수행하는 기술

위 기술들을 웹 트래픽 로그 공격 유형 분류에 있어서 성능 비교 평가

웹 로그 데이터 전처리

- 대소문자 통일
- URL 인코딩 및 HTML이스케이프 문자 복원
  - 예) %20"→""(공백) , "<script>"→"<script>“
- NUL 문자,중복 데이터 제거

TF-IDF

TF와 IDF를 결합한 값

`TF` `Term Frequency`: 문서 내 단어 빈도

(한 문서 내에서 특정 단어 빈도 수 / 전체 단어 총 빈도수)

`IDF` `Inverse Document Frequency`: 문서 집합 내에서 단어의 상대적 빈도

log (단어 등장 문서 수 / 전체 문서 수)

scikit-learn의TFidfVectorizer 통해 TF-IDF 벡터화 진행

단어 단위에 초점 맞추어 벡터화 진행 → 로그 데이터 단어들을 TF-IDF 가중치로 변환해 특징 추출 → 로그 데이터에서 특정 용어/패턴 중요도 파악 가능

SVD(Singular Value Decomposition)

TF-IDF 적용한 고차원 데이터를 저차원으로 변환 (차원 축소)

→ 데이터의 중요한 feature 추출 → AutoML, 딥러닝 모델의 연산 복잡도 낮추고 학습 속도 높임

NLTK 활용 ALBERT

NLTK로 웹 방화벽 로그의 공격 payload 데이터 전처리

- 토큰화
- 어근 추출
- 불용어 제거

전처리된 payload 데이터로 ALBERT 모델로 공격 패턴 분류

결과

- **AutoML:** 높은 정확도
- **TF-IDF+SVD:** 전처리와 함께 사용했을 때 좋은 성능, 속도도 빠름
- **ALBERT**: 가장 높은 정확도 (문맥 이해 강점), 그러나 자원 소모 크고 속도 느림
- **CNN/RNN**: 빠르지만 정확도 낮음 → 실시간 대응에 적합

⇒ BERT 기반 모델((그중에서도 ALBERT 모델)이 가장 적합하지만, 자원 소모를 줄이기 위해 **TF-IDF+SVD를 사용한 AutoML**이 실용적 대안으로 보임

앞으로 필요한 것

- ALBERT 최적화
- TF-IDF+SVD 성능 향상
- 더 다양한 로그 데이터 수집 및 정교한 전처리