**삼성DS Security Expert과정 PBL 과제 수행 계획서**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 과제명 | | AI 기반 스팸/악성 메일 탐지 | | | 시작일자 | 종료일자 |
| 2023-09-04 | 2023-12-01 |
| PBL 문제  정의 | | *※ pbl 문제의 배경, 특징, 목표, 기대효과 등을 간략하게 기재*  삼성전자 사내망으로 침투를 시도하는 스팸/악성 메일은 하루 20만개 수준이며 점점 증가 추세에 있어 보안 위험성이 증대되고 있다. 하지만 현재는 전통적인 스팸/악성 메일 필터링 방식이나 관계사의 솔루션에 의존하여 1차 방어를 하고 있어 이로 인한 별도 예외 프로그램 사용 등 다양한 행정적 절차나 결재들이 발생하여 임직원 업무 효율 저하를 발생시키고 있다. 또한 최근에는 인공지능 언어 모델이 발전함에 따라 인공지능 언어 모델이 만든 악성 메일 등이 스팸/악성메일 예방을 방해하고 있다. 그래서 기존 Rule based를 대표하는 철자와 문법 오류 등으로 필터링하던 방식을 회피하게 만들어진 스팸/악성 메일을 탐지할 필요성이 대두되고 있다. 따라서 AI를 기반으로 스팸/악성 메일 탐지 효율화를 통해 임직원 업무 효율성을 증대할 수 있도록 개선안을 제안한다. | | | | |
| PBL 문제  해결 방법 | | *※ PBL 문제를 해결하기 위한 예상 연구 방법 제시*   1. 스팸/악성 메일의 제목, 본문 내용을 NLP(Natural Language Processing) doc2vec의 PV-DM(word vector 생성 필요)을 이용하여 word embedding하여 dense representation을 선정 2. 선정된 Feature Representation을 검토하고 데이터 정제(EDA) 수행 3. 정제된 Feature Representation을 통해 머신 러닝 모델 학습 4. 테스트 데이터셋으로 머신러닝 모델을 테스트하여 결과를 검증 | | | | |
| PBL  기대 효과 | | *※ PBL 문제 해결 시 예상 산출물 및 기대 효과*   * 1. 기존 Rule-Based 스팸/악성 메일 필터링에서 AI 기반 필터링 기능을 추가하여 학습 샘플 데이터 기반으로 수신된 메일의 제목 및 본문을 탐색하여 스팸/악성 메일을 탐지하고 주의사항을 사용자에게 알려 사용자가 메일을 오픈하기 전 주의할 수 있도록 하여 사내 보안을 강화하는 효과를 기대할 수 있다.   2. Or 탐지된 스팸/악성 메일을 샌드박스 환경에서 오픈하여 문제 여부를 실제 테스트한 뒤 문제가 없으면 정상처리 문제가 없으면 차단 후 데이터베이스로 쌓아 사내 보안을 강화하는 효과를 기대할 수 있다.   3. 임직원이 업무를 수행함에 있어서 스팸/악성 메일을 거르는 데에 스트레스 받지 않고 업무에만 집중하여 업무효율성을 높일 수 있다.   4. 위 학습모델은 추후 사내망에서 사외망으로 나가는 메일을 검증하는 데에도 사용할 수 있을 것으로 기대된다. | | | | |
| PBL  예상 추진 일정 | | 9월 | 기존 연구, 논문 학습  Data preprocessing 방법 연구 및 학습  Data sets 확보 | | | |
| 10월 | Data preprocessing | | | |
| 11월 | Model Training / Model Selection / Fine Tuning | | | |
| 12월 | PBL 과제 결과 발표 | | | |
| 연구  지도 | 교수  (필수) | 성명 : 이상근 교수 | | 연락처 : sangkyun@korea.ac.kr | | |
| 성명 : | | Email : | | |
| 조교  (선택) | 성명: 권준형 | | 연락처 : imjulian@korea.ac.kr | | |
| 성명: 남현경 | | Email : greena@korea.ac.kr | | |

**PBL 진행사항(~ 9/7)**

1. Knox Portal(삼성전자 사내포탈) 삼성 SDS 보안담당자 확인 및 사내 필터링 방법 문의

: Knox Portal e-mail 필터링은 브로드컴社의 SMG Solution 을 사용하고 있음

링크: [Secure your email communications (broadcom.com)](https://www.broadcom.com/products/cybersecurity/email/gateway)

첨부: 하단 캡쳐본

1. PBL 프로젝트에서 진행할 Deep Learning 기반 방식은 기존 Rule Based 방식 혹은 ML 방식에 비해 새로운 유형이나 방식의 스팸/악성 메일이 유입되었을 때 어느정도 대처가 가능할 것으로 기대됨.
2. 기존 e-mail 스팸/악성 e-mail 필터링 업체에서는 왜 Deep Learning 을 도입하지 않았는지 그 이유를 알아볼 필요가 있을 것 같음.
3. SDS 담당자께 Knox Portal에 실제 유입 e-mail 데이터셋을 확보할 수 있을지 문의

(개인 메일로 오는 것들이 많을 것으로 확보가 어려울 것으로 판단됨)

1. 사내 실제 데이터 확보가 어렵다면 영문 e-mail 데이터셋 or 한글 e-mail 데이터셋을 사용할지 결정 필요.
2. 한글의 경우에는 데이터셋 자체가 확보하기 어려운 상황. 한국어의 경우 자립/의존 형태소가 존재하여 토큰화 할 때 영어처럼 간단하지 않아 Sentence Segmentation 어려움. KSS(Korean Sentence Segmentation)이 있는데 이를 활용하여 진행 예정.
3. 학습할 논문 선정

Spam email detection using deep learning techniques (인용 72회)

Machine learning for email spam filtering: review, approaches and open research problems(인용 420회)





