

# 역할분담(유효곤)

---

- 추가 실험
  - 난독화 종류 분포
  - 난독화에서 탐지율과 속도
  - Git 코드 합치고 Demo 찍기
- 이론적 내용
  - Why 딥러닝
  - Why CNN ( Why 저런 모형)
- 발표자료
  - 자료 변경
    - 최대한 말을 해야하는 정보를 ppt에 추가하는 것이 좋음.
    - 폰트 통일 + 폰트 사이즈가 3가지 정도만 잇 것
    - 14page - 각각이 어떤 난독화가 적용되었는 지 표시하고
    - 각 section을 요약하는 + 본인들의 성과를 알려주는 하는
    - 전문용어를 써서 설명하는 것들이
    - 말을 줄일 필요가 있을 듯 (말을 해야하는 핵심 포인트는 슬라이드에 있게)
  - 발표 코멘트
    - 강조하고 싶은 부분에서 말을 많이하는 것이 좋음.
    - Block Parsing 시간을 너무 많이 사용함.

# 역할분담(조규환)

---

- 추가 실험
  - git 코드 정리
  - 데이터 정규화에서 '빼기'
  - 확률로 0.5로 하였다면, ROC를 그리고 이를 튜닝하는 과정을 거쳤는 지? (개선으로 남기는 것도 좋을 듯)
  - 토큰의 개수를 512, 1024, 2048, 3072, 4096개부터 늘려가며, 보았을 때 어떤 값이 가장 좋았다.
  - 토큰 뽑는 걸 초반부, 중반부, 후반부로 했을 때 어떤 성능 차이가 있는 지? 후반부가 잘 안되는 이유는 무엇인지?
  - 발표에 토큰 중복제거 100개 (100개를 threshold로 잡은 이유)

# 역할분담(정호진)

---

- 시각화 추가
  - Projection : legend 사용 + 색상 변경
- 이론적 내용
  - f1, tfidf, recall, precision, word2vec, (test accuracy vs validation accuracy) 정의
  - Why Ours is good? (especially on recall)
  - How 개선
  - Image reference
  - Early stopping timing
  - Why 앞에 애들이 좋은가? (현상 X 설명 O)
  - Why 2048개?

# 모두가 볼것

---

- 애네들이 잘 이해하고 있는 지 보는 게 제일 크다.
- Why Ours is good? (especially on recall)
- 적용해보았더니 되더라. 를 제일 싫어함

# 이상근 교수님 코멘트

- 최대한 말을 해야하는 정보를 ppt에 추가하는 것이 좋음.
- 14페이지 : 각각이 어떤 난독화가 적용되었는 지 표시하고
- 강조하고 싶은 부분에서 말을 많이하는 것이 좋음.
- 폰트 통일 + 폰트 사이즈가 3가지 정도만 잇 것
- 왜 딥러닝 기법을 사용하는 지
- 적용해보았더니 되더라. 를 제일 싫어함
- 평가지표에 대한 이야기 (f1-score가 무엇인지, prec, recall)
- 난독화 종류 분포
- TFIDF가 무엇인지
- Projection : legend 사용 + 색상 변경
- recall이 높고 precision이 조금 낮은 데, 성공이유에 대해서 말을 해보아라.
- 개선을 한다면 어떤 것을 할 것인지?
- 난독화에서 탐지율과 속도
- 각 section을 요약하는 + 본인들의 성과를 알려주는 하는
- Script Block Parsing 시간을 너무 많이 사용함.
- 필요한 부분만 이야기
- Word2Vec 정의가 잘못 됨. 이 근거로 하는, NLP에서의 철학이 Distribution이.... 단어의 context 정보를 벡터화 하는 것. 또한, Context
- Word2Vec reference
- 구조의 이유 ( 질문 거리 )
- 답변을 준비하고 있다가,
- early stopping + overfit되는 것을 체크 했는 지?
- 확률값 0.5인가? logit 값...
- 확률로 0.5로 하였다면, ROC를 그리고 이를 튜닝하는 과정을 거쳤는 지? (개선으로 남기는 것도 좋을 듯)
- 전문용어를 써서 설명하는 것들이
- 말을 줄일 필요가 있을 듯 (말을 해야하는 핵심 포인트는 슬라이드에 있게)
- \*\* 애네들이 잘 이해하고 있는 지 보는 게 제일 크다.\*\*
- test accuracy와 validation accuracy의 차이 overfitting
- validation accuracy를 계속 상승시키려고 했기에
- Decision Boundry가 non-linear

# 이상진 교수님 코멘트 (원본)

---

- 1. 피피티를 문장형이 아니라 항목단위로 바꾸기
- 2. 512, 1024, 2048, 3072, 4096개부터 늘려가며, 보았을 때 어떤 값이 가장 좋았다.
- 3. 단어의 개수를 단순히 2048이 아니라 범위를 보여주며, 어떻게 좋은지, 왜 초기 단어를 사용해야 하는가?
- -> 설명으로 악성 스크립트들이 앞부분에 특징이 있는가? 초기 단어가 잘 되는 설명이 필요
- (악성 스크립트들이 길이가 길수 없어서, 초반에 행위를 한다던지, 후반부에는 정리하는 행위를 한다던지...)
- 4. 초반부, 중반부, 후반부로 했을 때 어떤 성능 차이가 있는 지? 후반부가 잘 안되는 이유는 무엇인지?