

[캡스톤디자인 중간보고서]

■ 연구과제

과제명 (작품명)	악성 댓글 분류 모델들의 성능 향상을 위한 방법 제시	참여학기	2021 년 1 학기
--------------	-------------------------------	------	-------------

■ 강좌정보

과목명	데이터분석 캡스톤디자인	학수번호	SWCON321-00
과제기간	2021 년 3 월 5 일 ~ 2021 년 6 월 18 일	학점	3

■ 팀구성

팀명	김효준		팀구성 총인원	1명
구분	성명	학번	소속학과	학년
대표학생	김효준	2016100946	산업경영공학과	4
참여학생				

■ 지도교수 확인

지도교수	성명	이대호	직급	전임교수
	소속학과	컴퓨터공학과	지도교수 확인	성명 : (인)

■ 붙임

[첨부1] 과제 중간보고서

본 팀은 과제를 성실히 수행하고 제반 의무를 이해하여 이에 따른 결과보고서를 제출합니다.

일자 : 2021 년 4 월 30 일

신청자(또는 팀 대표) 김효준 (인)

[캡스톤디자인 과제 중간보고서]

과제명	악성 댓글 분류 모델들의 성능 향상을 위한 방법 제시
<div>1. 과제 개요</div> <div>가. 과제 선정 배경 및 필요성</div> <div>- 최근, SNS가 활발해지면서 영화, 쇼핑, 뉴스 등 다양한 산업 분야에서 많은 익명의 리뷰들이 등장</div> <div>- 익명의 특성으로 인해 단순 욕설, 차별적인 말, 타인 비하 등 불쾌감을 주는 리뷰로 인해 피해받고 있는 실정</div> <div>- 현재, 네이버 AI 악플 감지기 서비스가 있지만, 반어적인 어투 또는 비꼬는 말 등의 여러 악성 리뷰를 판단하지 못하는 경우가 있기에 더 높은 정확도가 요구되는 모델 필요</div> <div>나. 과제 주요내용</div> <div>■ 데이터셋 선정</div> <div>1) Korean HateSpeech Dataset</div> <div>2) 욕설 감지 데이터셋3) AI 허브 인공지능 윤리 연구를 위한 비정형 텍스트 데이터셋</div> <div>■ 분석 방향</div> <div>1) 텍스트 전처리 - 불용어 제거</div> <div>- 형태소 분석 (KoNLPy 라이브러리 사용) : 명사, 형사 추출</div> <div>- 3가지 워드 임베딩</div> <div>① 빈도 수를 통한 워드임베딩 ② Word2Vec 모델을 통한 워드임베딩 ③ fasttext 모델을 통한 워드임베딩</div> <div>2) 모델링</div> <div>- RandomForest, DNN, BiLSTM 등 모델 구축</div> <div>- 텍스트에 있어서의 Data Argumentation을 통한 모델 성능 개선</div> <div>다. 최종결과물의 목표</div> <div>- 텍스트의 전처리하는 과정 및 모델 아키텍처를 구체적으로 학습하고자 함</div> <div>- 기존 논문에서 사용되는 모델을 구축하여 데이터셋에 적용시켜 성능을 도출한 후, Data Argumentation을 통해 성능 개선을 하고자 함</div>	
<div>2. 과제 수행방법</div> <div>가. 과제수행을 위한 도구적 방법 (활용 장비, 조사 방법론 등)</div> <div>- Google Colab Pro 서버를 이용</div> <div>- 관련 모델은 논문으로부터 아키텍처 공부</div> <div>나. 과제수행 계획</div> <div>- 텍스트 전처리 세부 공부 (라이브러리 위주)</div> <div>1. 형태소 분석기 2. 정수 인코딩 3. Word2Vec 4. FastText</div> <div>- 전처리 적용</div> <div>- 텍스트 예측 모델 공부 (논문 위주)</div> <div>- 모델링 적용</div> <div>- 모델 검증 및 성능 개선</div>	

### 3. 진행내용

#### 가. 과제진행 내용

- 악플 이진 분류 task를 집중하여 전반적인 텍스트 전처리 기법, 모델링 기법 등을 공부하였음

#### ■ 텍스트 전처리

- 형태소 분석기 중 가장 빠른 속도의 분석기를 선택하여 적용
- 불용어 제거

#### ■ 정수 인코딩

1. Bag of words (CountVectorizer) 기반의 Embedding layer 사용
2. 악플 리뷰 6백만건을 직접 학습시킨 Word2Vec 임베딩 사용

#### ■ 모델링

- 다음 사진의 기본 baseline으로 도출하고, 이를 개선하고자 함

	모델명	test 정확도
0	DNN	0.642614
1	LSTM	0.829670
2	LSTM_2layer	0.640797
3	Bi-LSTM	0.640797
4	Bi-LSTM-2	0.640797
5	1D-CNN	0.843407

- 개선 위해 파라미터 조정해가며 성능 비교( 휴리스틱 기법)
- Word2Vec 임베딩 벡터를 입력값으로 사용하여 성능 개선

#### 나. 진행내용의 주요특징 및 설명

- loss가 nan으로 개선되지 않는 모델 (2층 이상의 구조가 층이 쌓이면서 gradient vanishing 현상)

- ReLU 함수는 음의 영역이 미분값이 0 이므로 활성화 함수를 바꿔 해결

#### ■ Word2Vec 임베딩 벡터 사용

- 기존 DNN 모델의 경우, 0.64의 정확도에서 Word2Vec 임베딩 벡터를 입력값으로 사용해줌으로써, 모델의 정확도 개선을 향상시킬 수 있었다. (0.7798)
- 그러나, 임베딩 벡터는 이미 Embedding 되어 있는 형태로 LSTM, 1D-CNN 등 다른 모델에 적용하는데 있어서 차원이 안 맞는 실정
- 위의 문제를 개선하는 방향으로 향후 계획이 진행될 예정

### 4. 향후계획

순번	추진내용	5월	6월
1	모델링 공부 및 개선	■	
2	fasttext 인코딩 적용	■	
3	Data Argumentation 적용		■

팀 학생대표 성명 : \_\_\_\_\_ 김효준

김효준  
(인)