

2차 프로젝트_자연어처리

감정 기반 응답 챗봇

김영경

1. 프로젝트 배경 및 목적
2. 제작과정
3. 결론 및 느낀점

프로젝트 배경 및 목적

감성 대화 데이터를 활용한 챗봇 제작

감성 대화 데이터를 활용하여 감성적인 대화가 가능한 챗봇을 제작하고자 한다.

간접 대화를 통해 심리적 문제의 토로에 대한 접근성을 높이고, 우울증 등 부정적인 감정을 완화시키며, 궁극적으로 우울증 예방과 심리적 안정, 행복 증진 효과까지 기대해볼 수 있다.

진행 순서

DAY1,2 주제 탐색 및 데이터 확보
DAY3~7 모델 탐색 및 비교
DAY8~11 챗봇 제작, 발표자료 제작
DAY12~ 수정 보완

사용 툴

Python(vscode), Django(Pycharm)

사용 데이터

AI Hub, 감성 대화 말뭉치
[https://aihub.or.kr/aihubdata/data/view.do?currMenu=115
&topMenu=100&aihubDataSe=realm&dataSetSn=86](https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=86)

제작 과정

데이터 구성

약 5만 건의 train 데이터 & 6640건의 valid 데이터
텍스트 형식의 코퍼스 데이터 (60가지 감정을 개인 페르소나에 매핑한 감정 상태별 구어 텍스트 대화 데이터)

컬럼	내용
연령	청소년, 청년, 장년, 노년
성별	남성, 여성
상황키워드	진로,취업,직장 / 학교폭력,따돌림 / 대인관계 / 가족관계 / 재정 등등
신체질환	만성질환 유/무, 해당없음
감정_대분류	기쁨, 당황, 분노, 불안, 상처, 슬픔
감정_소분류	대분류별 9가지 세부 감정
사람문장1,2,3	감정 상태, 진전된 대화, 진전된 대화
시스템문장1,2,3	응답 호응, 응답 호응, 마무리 대화

종속변수

‘감정_대분류’ 컬럼을 라벨링.

0:기쁨, 1: 당황, 2: 분노, 3: 불안, 4: 상처, 5:슬픔

독립변수

‘사람문장’과 ‘시스템문장’만 사용

1 감정 분류 후 응답 출력 방식

CNN, RNN, LSTM을 이용한 감정 분류 모델을 제작한 후, 답변을 인출하는 챗봇을 제작하는 방식으로 진행했으나, 분류 모델의 정확도가 좋지 않았음.

```
- loss: 0.0000e+00 - acc: 0.1667 - val_loss: 0.0000e+00 - val_acc: 0.1671
```



감정 분류가 주 목적이 아니라, 감성적인 대화가 가능한 챗봇 구축이 목적이므로
감정 분류 모델보다는 맥락에 맞는 문장 생성 모델 구축으로 방향을 전환함.

2 BERT 모델을 사용한 문장 생성 모델 사용

BERT

트랜스포머를 이용해 레이블이 없는 텍스트 데이터로 사전 훈련된 언어 모델.

사전 훈련 모델을 가져와서 파인 튜닝 과정을 거쳐 적합한 모델로 추가 훈련시켜서 사용.

총 3개의 임베딩 층을 사용 (워드 + 포지션 + 세그먼트 임베딩) → 단어, 위치정보, 문장 → 맥락 파악 가능

SBERT

BERT로부터 문장 임베딩을 얻을 수 있는 센텐스버트.

학습법 1) NLI(문장 쌍 분류 테스트): 두 문장의 관계 맞추기 2) STS(문장 쌍 회귀 테스트): 두 문장의 의미적 유사성 구하기
이미 다양하게 학습되어진 모델들이 많아서, 본인의 데이터에 잘 맞는 모델을 찾아서 사용하면 되는 편리함.

SBERT 모델들의 출력값을 비교한 후 결정.

2 BERT 모델을 사용한 문장 생성 모델 사용

- ① sentence-transformers/xlm-r-100langs-bert-base-nli-stsb-mean-tokens
- ② jhgan/ko-sroberta-multitask
- ③ jhgan/ko-sbert-sts
- ④ ddbokki/klue-roberta-base-nli-sts-ko-en

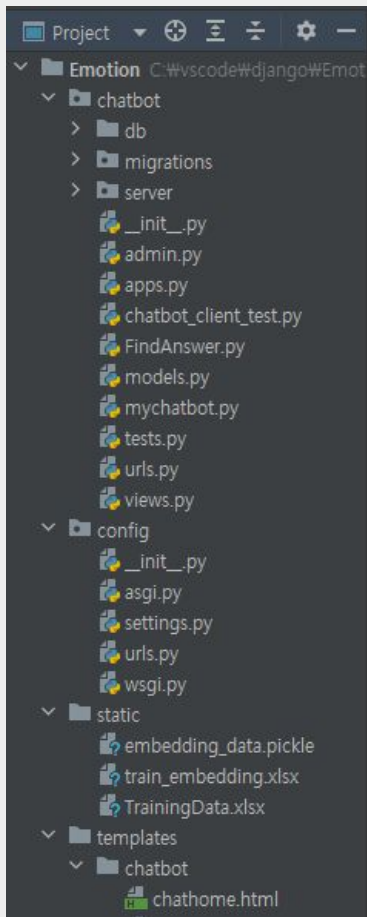
```
test = [
    '요즘 공부가 잘되어서 기분이 좋아',
    '죽고싶지만 떡볶이는 먹고싶어',
    '일하는 중인데 너무 졸려서 집에 가고싶어',
    '친구가 커피에 고추가루를 넣어먹는데 대체 왜 그러는지 정말 당황스러워.'
]
```

① 공부법을 찾아 기분이 좋으신가 봐요. 공부법이 마음에 드나요?
남편과의 사별로 많이 슬프시겠어요.
아픈 몸으로 일하시느라 고생이 많으세요.
시트에 커피를 쏟으셨다니 정말 화나시겠어요.

② 공부가 잘 된다니 정말 좋으시겠어요. 특별한 이유가 있나요?
죽고 싶은 마음이 드실 정도로 마음이 아프시군요.
업무가 지연되었나요?
어떤 일로 이렇게 화가 나셨을까요?

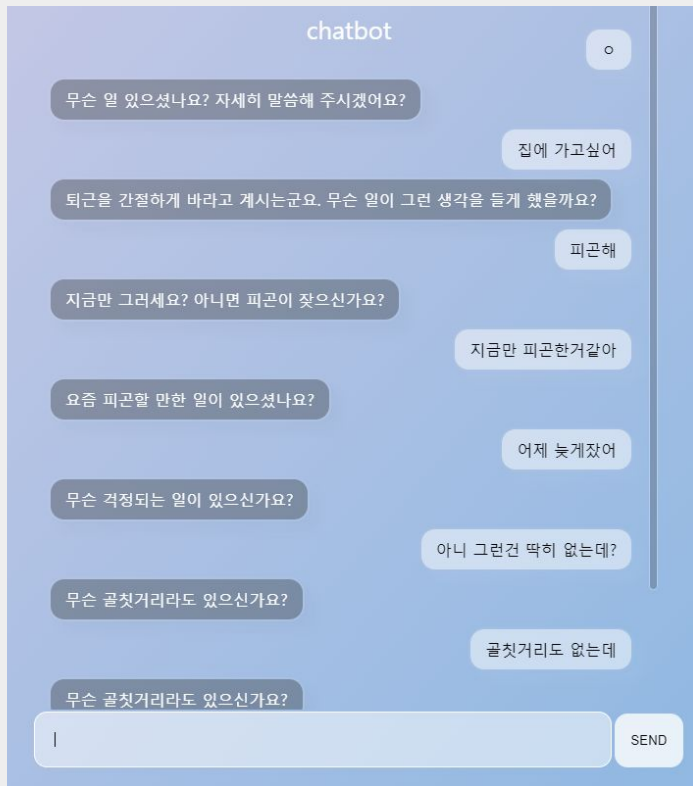
③ 공부가 잘 된다니 정말 좋으시겠어요. 특별한 이유가 있나요?
그래서 기분이 어떠세요?
잠이 와서 공부가 잘 안 되어 걱정하시군요.
식사 대신에 커피를 먹어서 큰일이 날까 봐 불안하겠어요.

④ 공부가 잘 된다니 정말 좋으시겠어요. 특별한 이유가 있나요?
죽고 싶다는 극단적인 생각을 하시게 된 이유가 있나요? 무슨일이신지 말씀해주세요.
회사 생활이 무척 피곤하시군요.
친구와 어떤 일이 있으셨던 건가요? 저에게 알려주실 수 있나요?



- Emotion 프로젝트, chatbot 앱 생성
- db, server, static 디렉토리 생성
 - ▶ static / 감성 대화 데이터, 임베딩 데이터(데이터의 임베딩값만 따로 저장한 파일)
- 챗봇 구동 py 파일
 - ▶ FindAnswer.py : 답변 검색 클래스 (class FindAnswer)
 - ▶ mychatbot.py : 데이터 로드 & 최종 답변 출력 함수 (def chatanswer)
 - ▶ views.py : home, query 함수 정의 (def query)
- 챗봇 UI html 파일
 - ▶ chathome.html

'**klue-roberta-base-nli-sts-ko-en**'로 제작한 챗봇
한 턴의 대화만 했을 때는 공감 + 다음 대화 유도 기능까지
잘 수행하는 것 같아 선택했지만, 대화를 이어나가려하니
계속 질문으로 답이 와서 추궁 당하는 기분이 든다.

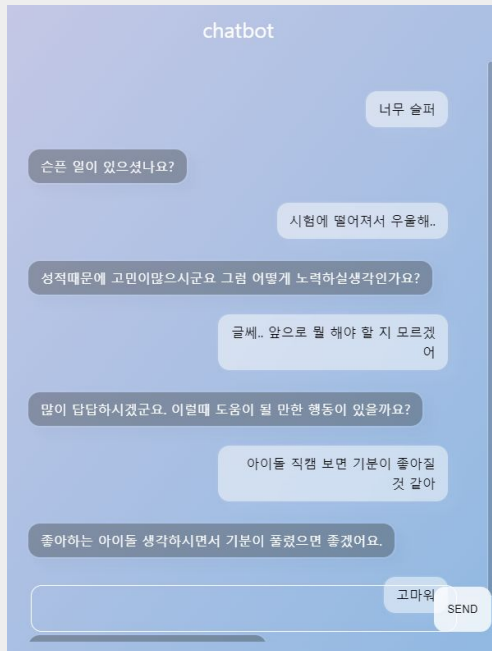


대화 데이터 추가

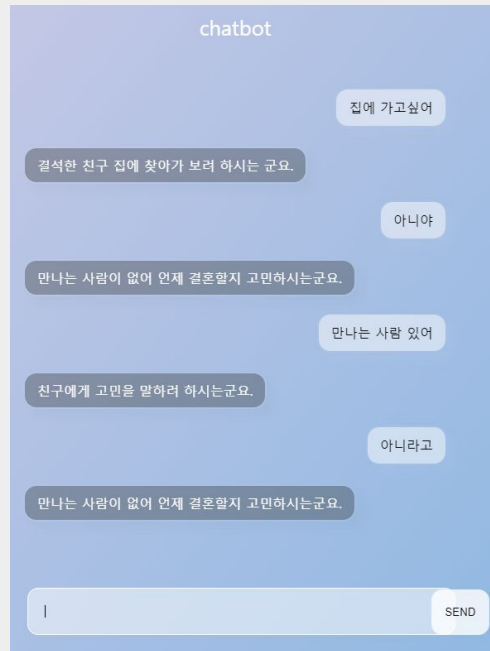
사람문장1과 시스템문장1만을
사용했었는데, 2,3 세트 추가.

결과

감정에 대한 대화는 잘 이어지나,
일상 대화는 매끄럽지 않은 대화로 이어짐.



감정에 대한 대화



일상 대화

결론 및 느낀 점

결론

SBERT 모델과 감성 대화 데이터셋을 이용한 문장 생성 모델을 만들어 챗봇에 적용했다. 사용한 데이터셋이 감정적인 대화에 집중된 데이터로 이루어져있어서인지, 일상 대화보다는 감성 대화인 경우에 챗봇과 대화가 잘 이어지는 모습을 보였다.

느낀 점

기본 모델에 직접 추가 학습을 시켜서 만든 모델을 사용했다면 좋았겠지만 시간 문제로 이미 학습되어진 SBERT 모델을 활용한 점이 아쉽게 느껴진다. 추후 직접 SBERT 모델을 만드는 과정까지 추가해보고싶다.

또한 이번 프로젝트에서는 주어진 데이터셋의 답변 값을 기반으로 응답을 하는 챗봇을 구현했지만, 페르소나를 가진 챗봇 모델을 구축할 수 있다면 더욱 사람과 대화하는 느낌을 받을 수 있을 것 같다.

