



LLM_1016

감정예측 및 노래 추천 챗봇

(빅데이터 9기 전동인)

요약

- 서비스

AI hub의 '감성 대화 말뭉치 데이터'를 활용하여 koelectra모델을 학습시켜 챗봇에 사용할 모델을 구축하고 사용자의 대화를 입력 받아 3번의 대화 이후 감정을 예측하여 미리 라벨링 해둔 음악을 추천해주는 서비스

- 대상

- 음악에 대한 구체적인 선호가 없이 대부분 좋아하여 음악을 선택하기 애매한 사람
- 본인의 감정이 불안정하여 대화와 음악을 통해 위로가 필요한 사람

실행 예시

감정 예측 챗봇

나: 행복해

챗봇: 무슨 일이 있었나요?

나: 너무 행복해

챗봇: 어떤 감정인지 더 설명해 주실 수 있나요?

나: 기분이 좋아

챗봇: 감정: 분노

오늘 기분이 '분노'하신가요? 이런 노래를 추천드려요: [Calm Down - Rema](#), [Let It Be - The Beatles](#)

|대화를 입력하세요

보내기

모델

- monologg/koelectra-base-v3-discriminator

- ElectraTokenizer를 활용하여 텍스트를 토큰아이징
- ElectraForSequenceClassification를 활용하여 모델 훈련. 훈련, 테스트 비율 8:2로 데이터를 나누어 학습시킴. 데이터를 확인하면 사람 대화가 2묶음에 각 대화가 어떤 감정인지 6가지로 라벨링이 되어 있는데 대화를 X, 감정 상태를 y로 학습
- 데이터가 10000개가 넘어서 모델 훈련 시 소요 시간 관계상 데이터에서 랜덤하게 20%만 추출하여 훈련

```
training_args = TrainingArguments(
    output_dir="./results",
    evaluation_strategy="epoch",
    per_device_train_batch_size=32,
    per_device_eval_batch_size=32,
    num_train_epochs=2,
    weight_decay=0.01,
```

웹 프레임워크

- Flask
 - 플라스크를 통하여 위에서 학습시킨 모델을 받아와서 3번의 대화를 거친 후 감정상태를 예측하여 미리 라벨링 해둔 노래와 매칭하는 챗봇 구축
 - 감정은 학습 때와 같게 6가지로 나누었고 노래는 각 카테고리 별로 2곡씩 변수에 저장.
 - 챗봇의 대화를 미리 저장해두고 사용자가 자신의 감정을 이야기하면 3번까지 감정을 이야기하도록 계속 유도

수정

- LLM 모델 추가
 - 학습 데이터가 부족하고 감정 라벨이 긍정에 해당하는게 한 가지 밖에 없어서 긍정의 문장을 입력했을때 분노로 출력되는 현상 발견
 - 'nlptown/bert-base-multilingual-uncased-sentiment' 모델을 koelectra 모델을 사용하기 전 문장이 긍정인지 부정인지 구분하고 긍정일 때 기쁨으로, 긍정이 아니면 koelectra 모델을 거쳐서 최종 출력이 되도록 수정.
- 노래 링크 추가

- 사전에 저장해 둔 노래들에 대해 유튜브 링크를 같이 저장하여 노래가 추천되면 링크로 넘어갈 수 있게 작업

한계

- LLM모델을 추가하였으나 여전히 기쁨의 감정을 잘 구분하지 못함. 3번의 대화를 종합하여 긍정의 상태로 구분하는 것이 한계가 있는 것인지 의문. ⇒ 사전에 학습하지 못했던 80%의 데이터로 추가로 학습할 필요가 있음(해당 PC환경 기준 11시간 정도 소요 예정)
- 감정 상태에 따라 노래가 현재는 변수로 저장되어 있어서 방대한 음악을 활용할 수 없음, ⇒ API를 이용하여 감정에 대해 음악을 추천받거나 음악을 저장해둔 DB에 각 음악별로 감정을 라벨링하여 호출될 수 있게 개선
- 챗봇 역시 사용자의 3번의 답변 입력을 유도하기 위해 같은 말을 반복적으로 제시하는데 답변을 추가로 늘려 랜덤하게 제시.