



2022 KU SW 경진대회 일기 감정 기반 영화 추천 시스템

팀 소개



팀 명 macs

지도교수 김학수 교수님

구 성 원

이름	학번	학과	주요역할
정창해	201710304	수학과	감정 분석 모델 설계 및 구현
손소영	201711649	응용통계학과	감정 분석 모델 설계 및 구현
김예지	201814246	시스템생명공학과	감정 분석 모델 설계 및 데이터 라벨링
박인우	201512265	컴퓨터공학부	앱 개발 및 서버 개발

작품 소개

작 품 명 일기 감정 기반 영화 추천 시스템

개발기간 2022년 3월 - 11월 (총 8개월)

작품 요약 사용자가 일기를 작성하면, 일기를 분석하여 그날의 감정 분포를 보여주고, 더불어 그 감정과 유사한 영화를 추천해주는 iOS 기반의 앱

적용 기술 Python, PyTorch, Transformers, Swift, Combine, MVVM, Flask

작품구성 및 세부내용

프로젝트 소개 Introduction

개발 배경

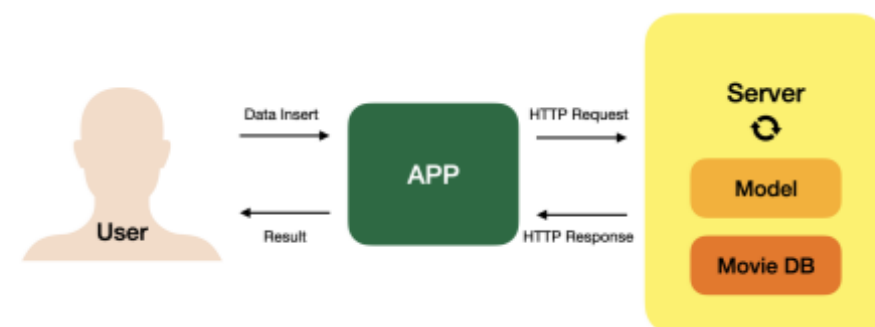
- 내면의 감정을 억압하는 게 일상적인 사회^[1]
→ 자신의 감정을 이해하고 표출할 줄 아는 것이 정신 건강에 이로움
- ➡ 일기의 감정을 분석하여 그 날의 감정을 이해할 수 있도록 하자!
- 감정을 표출하는 방법에는 일기 작성, 문화 콘텐츠 시청 등이 있음
- OTT 플랫폼이 활성화된 오늘날 영화 시청에 시간적 제약이 없어지면서 보다 많은 사람들이 여가 시간에 영화를 시청
- ➡ 일기의 감정을 분석하여 그 날의 감정에 맞는 영화를 추천해주자!

모델 학습 데이터셋

- KOTE (Korean Online That-gul Emotions) dataset
다양한 플랫폼에서 수집한 50,000개의 댓글을 44개의 감정으로 라벨링한 데이터셋

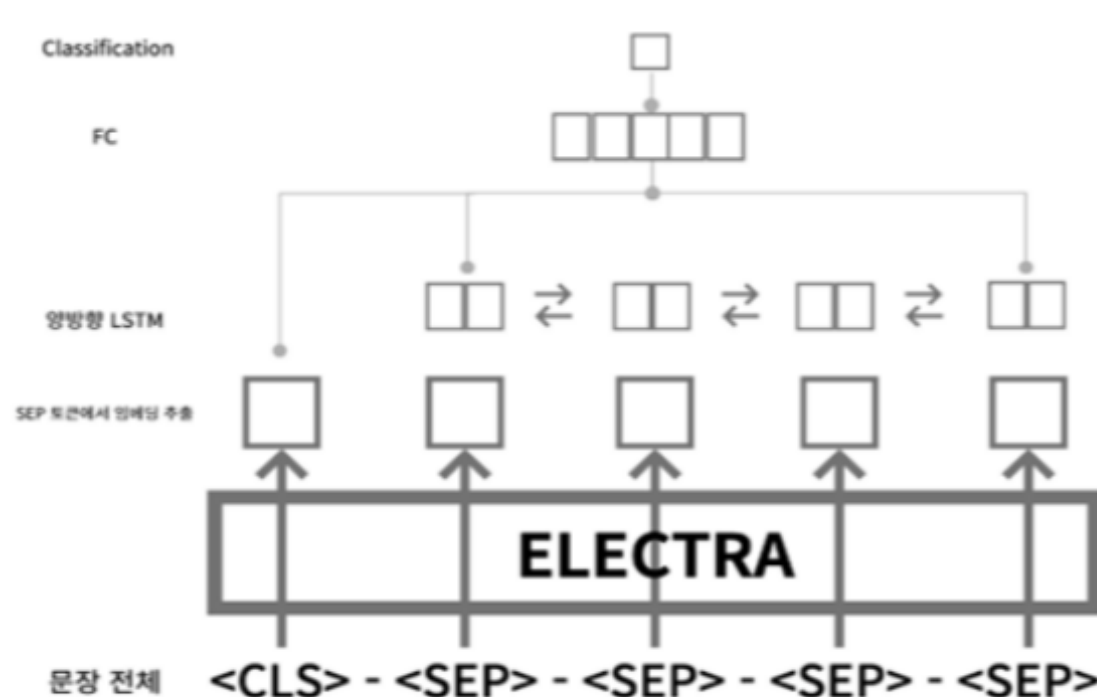
[‘불평/불만’, ‘환영/호의’, ‘감동/감탄’, ‘지극저극’, ‘고마움’, ‘슬픔’, ‘화남/분노’, ‘존경’, ‘기대감’, ‘우울함/무시함’, ‘안타까움/실망’, ‘비장함’, ‘의심/불신’, ‘뿌듯함’, ‘편안/쾌적’, ‘신기함/관심’, ‘아껴주는’, ‘부끄러움’, ‘공포/무서움’, ‘질망’, ‘한심함’, ‘역겨움/칭그러움’, ‘짜증’, ‘어이없음’, ‘없음’, ‘패배/자기혐오’, ‘귀찮음’, ‘힘들/지침’, ‘졸거움/신남’, ‘깨달음’, ‘죄책감’, ‘중오/혐오’, ‘호뭇함(귀여움/예쁨)’, ‘당황/난처’, ‘경악’, ‘부담/안 내림’, ‘서러움’, ‘재미없음’, ‘불쌍함/연민’, ‘놀람’, ‘행복’, ‘불안/걱정’, ‘기쁨’, ‘안심/신뢰’]

디자인



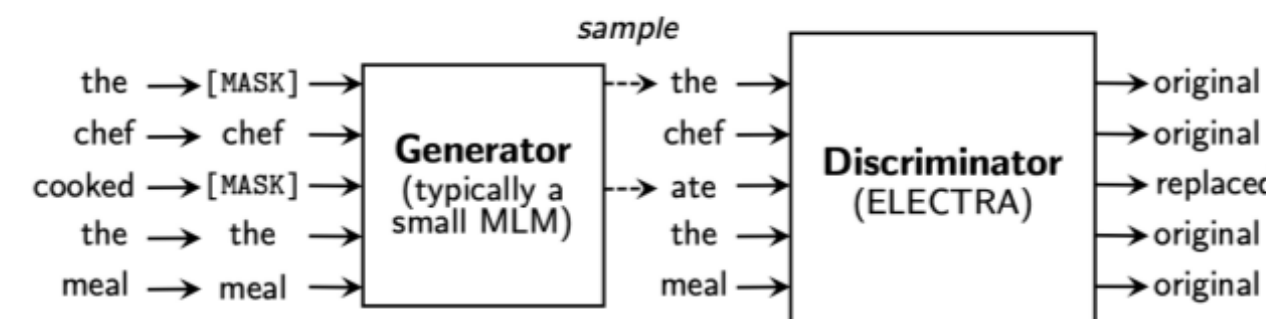
- 앱에서 사용자가 일기를 입력하면 서버에서 사전에 학습된 모델을 이용해 일기의 감정 벡터 반환
- 이를 미리 모델에 돌려서 db에 저장한 영화 벡터와 유사도를 계산하여 감정 확률과 영화 추천 결과를 반환

감정 분석 모델 Emotion Analysis Model



- 앱에 작성하는 일기의 특성 상 pre-train에 사용된 언어는 실생활에서 많이 사용되고, 온라인에서 많이 사용되는 단어들이어야 함
→ 한국어 댓글로 pre-train한 kcELECTRA 모델을 사용
- 데이터를 KSS(Korean Sentence Split)를 사용하여 문장 사이에 <SEP>토큰을 추가, <SEP> 토큰에서 추출한 임베딩을 bi-LSTM 레이어에 넣어 감정의 흐름을 반영
- <CLS>토큰, 첫 문장과 마지막 문장의 <SEP>토큰에서 계산된 loss를 합쳐 학습 진행
- 각 문장의 감정을 모델에 넣어 silver label을 생성한 후 직접 라벨링을 하여 gold label을 만들

ELECTRA



- PLM(pre-trained language model)로, BERT 파생 모델 중 하나
- BERT의 많은 계산량을 요구한다는 단점을 해결하기 위해 RTD (Replaced Token Detection) 기법 사용
RTD 기법: 실제 입력의 일부 토큰을 그럴싸한 가짜 토큰으로 바꾸고, 진위 여부를 맞추게 함 → 모든 토큰에 대해 학습하여 효율적이면서 효과적

bi-LSTM

- 일기의 문장의 흐름을 반영하기 위해 추가한 레이어
- sequential 데이터의 time step 간 양방향 장기 종속성을 학습

Cosine Annealing Scheduler

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min}) \left(1 + \cos \left(\frac{T_{\text{cur}}}{T_{\text{max}}} \pi \right) \right)$$

- local minimum에 빠지는 문제를 해결하기 위해 사용한 스케줄러
- learning rate가 cosine 함수를 따라서 eta_min까지 떨어졌다가 다시 초기 learning rate까지 올라감

결과 Result

모델	micro F1	차이
kcELECTRA	0.57	
kcELECTRA + bi-LSTM	0.62	+0.05
kcELECTRA + bi-LSTM + CosineAnnealingLR Scheduler	0.65	+0.03

- kcELECTRA+bi-LSTM 아키텍처를 적용하였을 때 kcELECTRA 단독 모델에 비해 성능이 5% 향상
- 여기에 스케줄러를 Cosine Annealing LR 스케줄러로 변경한 결과 성능이 추가로 3% 향상되어 총 8% 향상

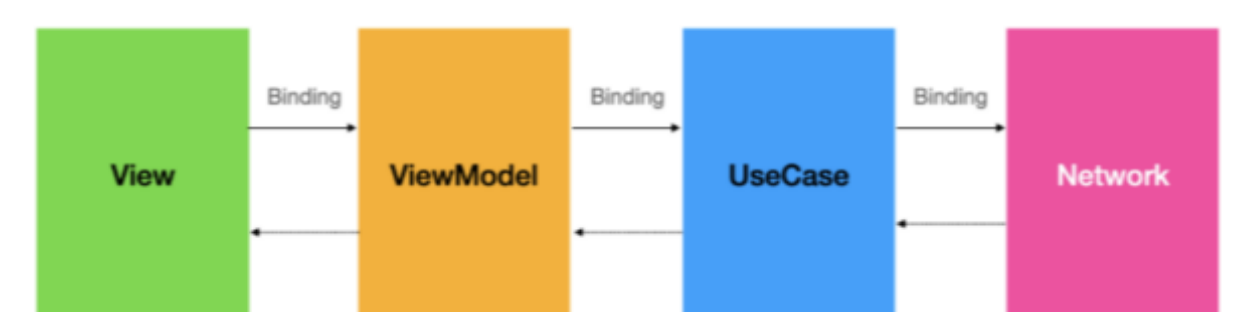
영화 추천 Movie Recommendation

코사인 유사도 (Cosine Similarity) 사용

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

- 두 벡터가 얼마나 유사한지 나타내는 척도
- 사용자가 작성한 일기와 영화 줄거리의 감정의 유사도를 측정 (영화 추천 시 사용자의 감정을 고려하였을 때의 추천 만족도가 고려하지 않았을 때에 비해 15.6% 증가했다는 연구 결과)^[2]
- 코사인 유사도를 바탕으로 일기의 감정과 제일 유사한 5개의 영화를 추천

iOS 앱 iOS App



앱

- MVVM 아키텍처 기반 시스템 구성
- 중간에 UseCase 계층 → 동일한 작업에 대한 중복 코드를 줄임
- Combine를 통한 데이터 Binding → 비동기적으로 데이터를 받아올 수 있는 형태

서버

- Flask를 통해 구현
- 앱에서 데이터를 요청했을 경우 그에 맞는 데이터를 확인하고 내보내도록 처리
- 요청은 REST API 형태의 HTTP Request로 구성

참고자료 Reference

- [1] [함영준의 마음 디톡스] (25) 마음의 소리나 감정을 잘 모른채 살아가는 현대인
<http://www.mindgil.com/news/articleView.html?idxno=69419>
[2] 감정 온톨로지 기반의 영화 추천 기법 (김국섭, 이석원, 2015)
<http://koreascience.or.kr/article/JAKO201531736561252.pdf>
Cosine Annealing Scheduler: <https://sanghyu.tistory.com/113>
학습 데이터: <https://github.com/searle-j/KOTE>
일기 테스트 데이터: 모두의 말뭉치 - 비출판물 말뭉치

기대 효과

하루하루 일기의 감정을 기록하는 “라이프 로깅”을 통해 사용자는 본인의 감정을 잘 이해할 수 있다.
그 날의 감정과 유사한 영화를 추천 받아 시청함으로써 본인의 감정을 표출하고, 보다 정신적으로 건강한 삶을 영위할 수 있다.

발전 계획

Hyperparameter tuning을 통한 모델 성능 개선

SW중심대학

KU 건국대학교
KONKUK UNIV.