

2022

A Poem Recommendation
Web Service Based on
Sentiment Analysis

감성 분석 기반의 시 추천 웹 서비스

홍익대학교 게임학부
유민경, 배병철

INDEX

1. 감성 분석 웹 서비스 제안
2. 웹 서비스 설계
3. KoBERT 감성 분석 모델 생성
4. 감성 분석 및 현대시 분류 방법
5. 결과
6. 결론 및 향후 과제

1. 웹 서비스 제안

A 트렌드

감성 분석(Sentiment Analysis)에 대한 관심 높아지며,
주어진 데이터셋으로 감성 분석 성능 향상을 위한 연구 활발히 진행

B 감성 분석의 활용

영화평, 뉴스 기사 등의 데이터에 적용하여 선호도 분석 및 챗봇의 자연어 생성 등의 응용에서 활용

C 문제점

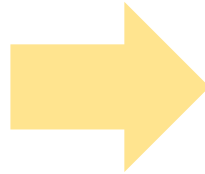
연구에 요구되는 많은 양의 데이터셋 (평균 1,000~10,000개 이상)
의미 있는 데이터셋 구축에 긴 시간 소요

1. 웹 서비스 제안

A 트렌드

B 감성 분석의 활용

C 문제점



“공개되어 다양한 용도로 사용되고
있는 NSMC(Naver Sentiment
Movie Corpus) 데이터 활용”



“사용자가 입력한 ‘자신의 하루’에
대한 감상평과 적합한 현대시를 제공
하는 웹 서비스 제안”

2. 웹 서비스 설계

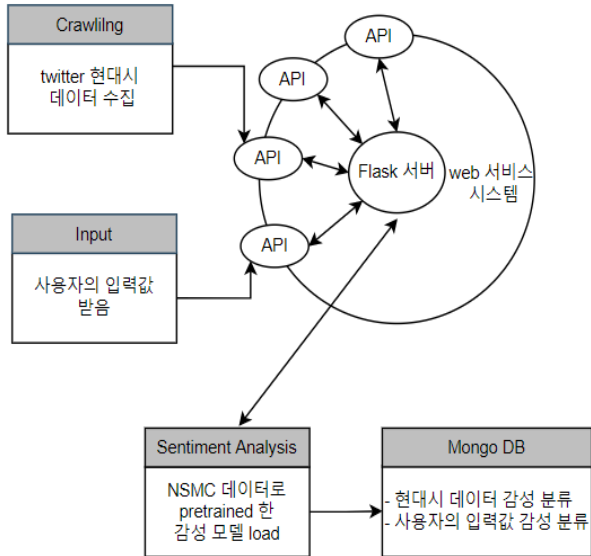


그림 1. 시스템 아키텍처

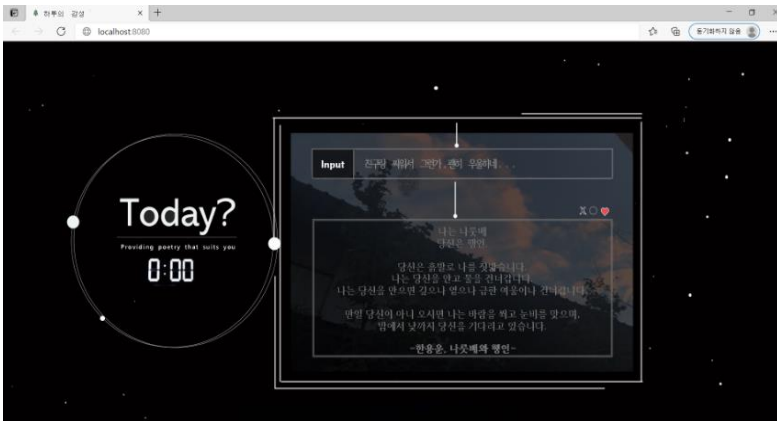


그림 2. 웹 서비스 제작 화면

KoBERT 감성 분석 모델

사용자가 입력한 자신의 하루와 크롤링한 시 데이터를 감성 분석할 때 사용

A Crawling

Twitter '현대시봇'의 트윗 데이터를 2021.04.01을 기준으로 3,000개의 현대시 데이터 수집

B MongoDB

수집한 현대시 데이터 저장
사용자 입력 값 서버에서 받아와 저장

C Input/Output

사용자의 '자신의 하루', 일기 형식의 문장을 받음
입력 값 감성 분석 모델 아키텍처 불러와 분석 후 분류된 시 출력

D Flask

웹 서비스 배포

3. KoBERT 감성 분석 모델 생성

A 연구 개요

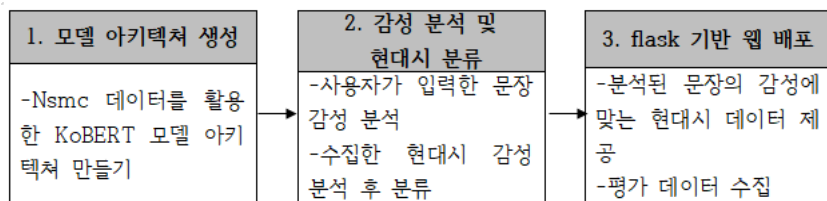


그림 1. 연구 개요 3단계

B 세부 설명

1. KoBERT에 맞는 입력값으로 NSMC 데이터 변환

- pretrained한 KoBERT Tokenizer를 import -> NSMC 데이터의 document 열에 해당하는 데이터를 [token input, segment input, mask input] 형태로 변환
- NSMC의 label 데이터는 별도의 변수 선언하여 각각의 값을 관리

2. KoBERT 학습

- 변환한 NSMC 데이터를 KoBERT의 train data로 활용
- 학습 안전성 높이기 위해 RAdam Optimizer를 이용

3. 감성 모델 확장

- NSMC는 긍정, 부정으로만 labeling 되어 있으므로 도메인 데이터와 pretrained한 모델에 dense layer 추가하여 새로운 layer를 쌓아 재학습
- epochs 조정하며 적절한 모델 찾기

4. 가중치 저장

- python 기반의 마이크로 웹 프레임워크인 Flask 기반의 웹사이트에서 모델 이용 시, 모델 아키텍처만 불러와 모델 로드하기 위함

4. 감성 분석 및 현대시 분류 방법

A 모델 로드와 Predict Score

- 사용자가 웹을 통해 데이터를 입력하면, 저장한 모델의 가중치를 불러와 모델을 로드하고 입력한 데이터의 predict score를 계산
- predict score는 0~1 사이의 값이 출력되며, predict score에 따른 감성 분석 기준은 표1과 같이 임의의 임계치에 따라 5가지로 정의

B 현대시 데이터 수집 및 저장

- python BeautifulSoup package 활용해 트위터 '현대시봇' 의 '트윗 및 답글' 란의 타겟 URL을 읽어 HTML을 쉽게 검색할 수 있는 상태로 변환
- 현대시 데이터가 속해 있는 태그를 찾아 텍스트 수집 후, 불용어(stop word)를 전처리 과정을 통해 제거
- KoBERT 감성 분석 모델로 현대시의 감성을 분류하고, 사전에 정의한 5가지의 감성 라이브러리 중 해당하는 곳에 저장

임계치 기준	감성
$0 \leq \text{predict score} < 0.2$	매우 나쁨
$0.2 \leq \text{predict score} < 0.4$	나쁨
$0.4 \leq \text{predict score} < 0.6$	보통
$0.6 \leq \text{predict score} < 0.8$	좋음
$0.8 \leq \text{predict score} \leq 1$	매우 좋음

표 1. 감성 분석 기준

4. 감성 분석 및 현대시 분류 방법

C 현대시 추천 기준

-분류된 현대시 데이터는 DB에 저장되어, 사용자의 입력 값에 대한 감성 분석을 마치고 웹페이지에 보여줄 때 사용자의 감성에 해당하는 라이브러리에서 랜덤으로 현대시 데이터를 추출해 감성 값과 함께 호출

D 데이터 수정

-사용자는 웹페이지에 호출된 감성 값과 현대시를 만족, 보통, 불만족으로 평가할 수 있고, 불만족일 경우 해당 현대시가 5가지의 감성 중 새롭게 분류되어야 할 감성을 택하게 함으로써 해당 현대시의 라이브러리를 수정

5. 결과

A 모델 성능 평가

overfitting 되지 않는 선에서 epochs=6까지 늘려 성능을 평가한 결과

**“89%
Accuracy”**

B 현대시 분류 결과

극성 NSMC 데이터 사용으로 인해 ‘보통’ 감성 값으로 분류된 시 많지 않았음

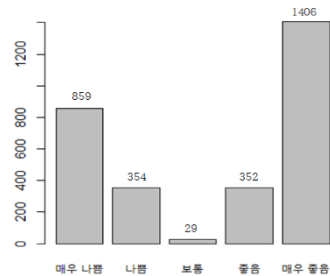


그림 1. 현대시 분류

C 만족도 및 기타 결과

인공지능 커뮤니티 사이트인 <한국 인공지능 커뮤니티> 등을 통해 일반 사용자들의 평가 수집

한용운의 <이별은 미의 창조>의 경우 ‘이별’과 ‘미’를 함께 나열함으로써 역설적인 표현으로 이별을 찬미하는 내용으로 5개의 라이브러리 중 ‘좋음’ 또는 ‘매우 좋음’ 쪽에 해당하는 것이 맞지만 ‘나쁨’에 분류되어 있었음. 의도와는 다른 값을 반환하였지만, 사용자는 이를 긍정적으로 평가하여 라이브러리가 수정되지 않았음.

수정된 라이브러리 중, 감동에 의한 ‘슬픔’이 ‘좋음’ 또는 ‘매우 좋음’으로 분류되어 있어 불만족한 경우도 있었다. 이처럼 원인과 결과에 따라 같은 감성이라도 다른 의미를 지닐 때 분류의 오류가 다수 발생했으며, 독자 역시 불만족하는 상황이 발생하기도 했다.

**“시의 비유, 상징적 표현 또는 역설적인 표현을 위한
별도의 분류 과정 필요”**

**“시의 다양한 표현 때문에 여러 가지로 해석될 수 있으므로
약간의 오류는 독자가 자신의 상황에 맞춰 해석하고 이해”**

**“감동으로 인한 슬픔, 죽음에 의한 슬픔 처럼 같은 감성이라도
다른 의미를 지닐 때 분류 오류 다수 발생”**

(단위: %)

평가 \ 데이터	감성	현대시
만족	88	72
보통	9	15
불만족	3	12

표 1. 각 데이터에 대한 만족도 평가

5. 결론 및 향후 과제

A 결론

만족도 평가(샘플 크기 = 200)
결과, 감성에 대한 만족도(88%)
가 현대시에 대한 만족도(72%)
보다 높게 나타났으며, 불만족
도 또한 감성 분석(3%)이 현대
시(12%)보다 매우 낮게 조사되
었음

감성 만족도

>

현대시 만족도

⇒ 상황에 맞춰 해석하고 이해하기 때문

향후 과제 및 계획 B

시의 표현(비유, 역설 등) 해석을 일반적으로
해석하기 위한 연구 필요

문맥 정보와 역설적 의미를 고려하여
감성 분석 시행 필요

**문맥 정보를 고려한 감성 분석
모델 생성 후 사이트 보완**

+

**수집한 시 데이터 셋 바탕으로
시 생성 모델 생성 후 서비스**