



카공 - AI/Data-Team

스터디 카페 추천 시스템

정지호, 장동훈, 이수빈, 양지인, 조현진, 문현정

작품 소개

선정배경

다양한 카페 매장에서 공부 환경이 조성되고 있지만, 공부에 적합한 환경을 갖추었는지를 알기란 쉽지 않다. 따라서 현 위치에서 사용자 맞춤 스타일로 공부에 적합한 가까운 카페를 추천함으로써 학생들이 시간과 비용을 절약할 수 있도록 하는 시스템이 필요하다.

작품개념

데이터베이스에 저장된 카페 이름, 평점, 평균가격, 공부환경점수, 카페 특색 및 리뷰 정보를 통해 사용자에게 적합한 카페를 추천 한다. 그리고 주기적으로 데이터베이스에 최신 정보를 추가로 저장하여 더 나은 추천을 할 수 있도록 한다.

역할 분담 및 진행 과정

1. 데이터 수집 파트

양지인, 조현진, 문현정
카페 정보 수집
리뷰 수집 및 라벨링

2. 모델 학습 파트

정지호, 장동훈, 이수빈
카페 정보 수집
모델 학습 및 자료 정리

1주차 :: 자료 조사

2주차 :: 개념설계

3주차 :: 전체 회의

4주차 :: 카페 데이터 수집

5주차 :: 시스템 및 모델 선정

6주차 :: 전처리 및 모델 학습

7주차 :: 학습 및 자료 정리

작품 내용

1. 데이터 수집

카페 정보와 리뷰데이터 수집

2. 전처리1

지도학습 라벨의 분포 확인(균등), 데이터 결측치 제거, 중복샘플 제거, 한글 공백 제외한 모든 문자제거, 빈샘플제거, 패딩(샘플의 길이 맞추기 60으로 통일), 불용어 정의, 정수인 코딩(단어 토큰화), 희귀단어빈도 확인

3. Lstm 감정분석

Train : Validation : Test = 7 : 1 : 2, Embedding dim=100, hidden unit size = 128, Activation = sigmoid & Loss = binary-crossentropy, Early stopping, train accuracy = 0.9807 & test accuracy = 0.8693

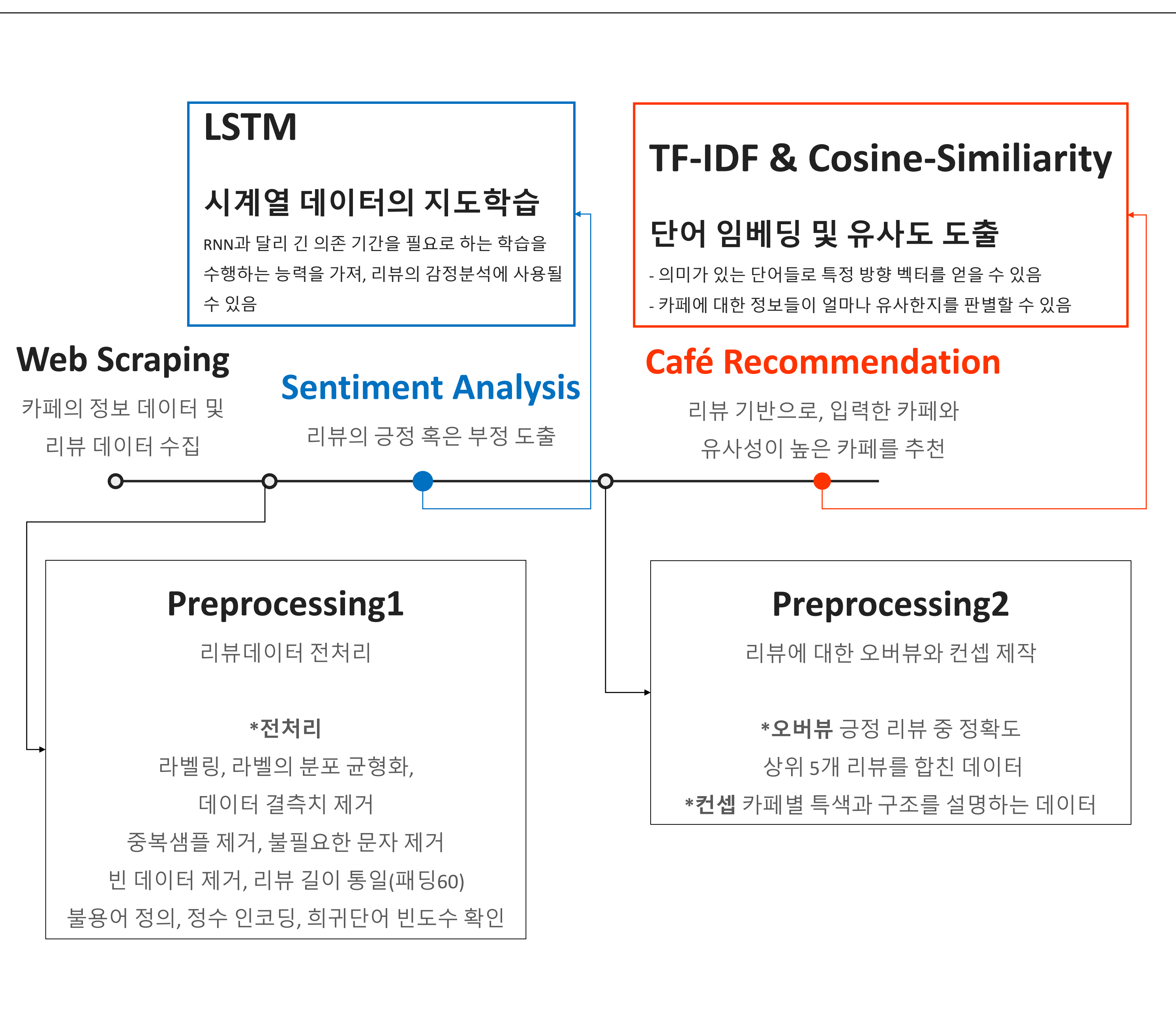
4. 전처리2

리뷰에 대한 오버뷰 및 컨셉 제작, 오버뷰란 lstm긍정으로 판별된 리뷰 중 정확도 상위 5개 리뷰를 합친 것, 컨셉이란 카페별 특색과 구조를 설명하는 문장(카페별 데이터 생성)

5. TF-IDF와 COS유사도로 카페추천

카페이름+오버뷰+컨셉 데이터를 활용, l2 정규화, 학습의 단위를 단어로 설정, 생성된 행렬로 코사인 유사도 계산, 유사성이 높은 카페들을 출력

시스템 설계



결과

```
In [45]: print(reco_top_similar_cafes('에이트', 5))
['에이트', '스타벅스 대전도안마점', '탐앤탐스 대전유성점', '파스쿠피 유성온천역점', '할리스 유성온천점']

In [46]: print(reco_top_similar_cafes('마을', 5))
['마을', '시크릿', '미디어 홍남대점', '소신', '할리스 유성온천점']

In [49]: print(reco_top_similar_cafes('미디어 홍남대점', 5))
['미디어 홍남대점', '공차 홍남대점', '카페소담', '루셈플레이스 홍남대점', '미디어 대전유성온천역점']
```

기대효과 및 고찰

사용자는 시간과 비용을 절감하며 효율적으로 공부환경을 선택할 수 있다. 총대 인근 위주인 82개 카페들의 데이터를 토대로 충남대학생을 위한 추천 시스템을 개발했지만, 향후 더 많은 데이터를 수집하여 전국에서 활용할 수 있는 시스템으로 발전시키고자 한다.