

자동 분류, 추천에 대한 기술 자료조사

문장 임베딩

TF-IDF(Term Frequency - Inverse Document Frequency) / BM25

BERT 계열 임베딩 모델

유사도 분석

추천 알고리즘

콘텐츠 기반 필터링

협업 필터링

참고

문장 임베딩

TF-IDF(Term Frequency - Inverse Document Frequency) / BM25

- 키워드 중심 비교
- 장점
 - 단어의 중요도를 수치화 가능
 - 불용어 걸러내기 쉬움
 - 구현이 간단
- 단점
 - 새로운 문서 추가 시 재계산 필요
 - 단어의 순서나 문맥 고려 어려움

Multi-Vector(CoBERT)

벡터DB

BERT 계열 임베딩 모델

- FastText
- BERT

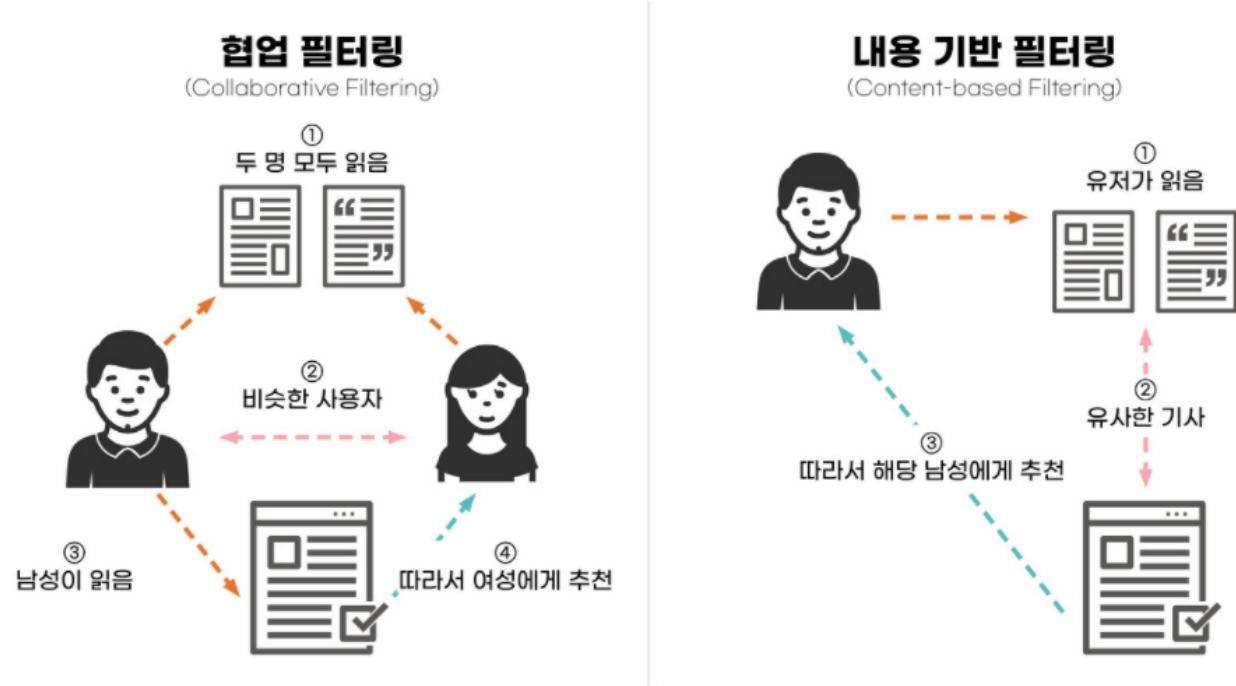
- Longformer, BigBird
- BERT(다국어), KoBERT(한국어 특화), Ko-SBERT(문장 임베딩 특화)

유사도 분석

- Elastic Search
- 벡터DB

추천 알고리즘

- 사용자 정보 혹은 콘텐츠의 정보를 기반으로 관련 콘텐츠를 제안하는 기술



콘텐츠 기반 필터링

- 콘텐츠에 관한 정보와 사용자의 선호도를 수집한 뒤 유사성에 따라 콘텐츠를 추천하는 방식
ex) 음악 콘텐츠를 선호하는 사용자가 있다면 주로 듣는 음악의 장르, 가수, 연도 등을 데이터화하고 분석하여 사용자가 선호할만한 또 다른 음악 콘텐츠를 추천
- 장점

- 사용자 행동 데이터가 적어도 콘텐츠 추천이 가능하다.
- 단점
 - 콘텐츠의 정보를 모두 함축하는 것에 어려움이 있다.
ex) 특정 캐릭터에 대한 추천 콘텐츠만 받고 싶어도 해당 콘텐츠의 정보가 많지 않으면 공통점이 있는 다른 캐릭터를 추천할 수 있다.

협업 필터링

- 동일 행동을 한 사용자들을 그룹으로 묶은 뒤 그들이 공통적으로 봤던 콘텐츠를 유사 행동 사용자에게 추천하는 방식
ex) 영화 콘텐츠의 경우 해당 영상을 봤던 사용자 그룹이 공통적으로 봤던 또다른 영화 콘텐츠를 사용자에게 추천
- 단점
 - 기존 데이터가 없는 신규 사용자의 경우 추천이 어렵다.
 - 정확한 추천이 이루어지기 위해서는 데이터가 쌓일 충분한 시간이 필요하다.

참고

어떻게 알았지 또 취향 저격! 추천 알고리즘의 원리가 궁금해
안녕하세요. 뉴스룸지기입니다. :) 동영상 사이트에서 좋아하는 영상을 보고 나면 이와 관련된 영상이나 자신의 취향에 꼭 들어맞는 영상을 추천받을 수 있는데요! 어떤 과정을 통해 엄선된 추천이 이루어지는 것일까요?
 <https://news.lginnote.com/967>



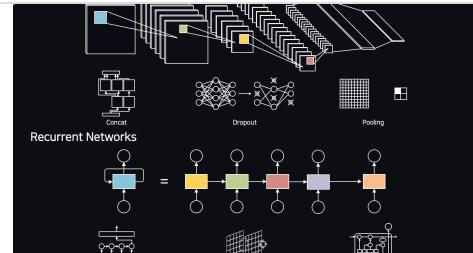
글쓰기 카테고리 추천모델 개발하기

안녕하세요. 당근마켓 머신러닝 엔지니어 Young입니다. 이번에 글쓰기 화면에 들어간 카테고리 추천 기능의 모델을 개발한 과정을 정리해 봤습니다.
 <https://medium.com/daangn/%EA%B8%80%EC%93%B0%EA%B8%B0-%EC%B9%B4%ED%85%8C%EA%B3%A0%EB%A6%AC-%EC%B6%94%EC%B2%9C%EB%AA%A8%EB%8D%B8-%EA%B0%9C%EB%B0%9C%ED%95%98%EA%B8%80-cbbcc43e1f7f>

TensorFlow를 활용한 네이버쇼핑의 상품 카테고리 자동 분류

"Auto-classification of NAVER Shopping Product Categories using TensorFlow"(English)

 <https://d2.naver.com/helloworld/1264836#>



어떤 추천시스템을 사용해야 할까? (1) - 협업 필터링 모델과 한계점

목차

d. <https://deepdaiv.oopy.io/articles/1>

Article

어떤 추천시스템을
사용해야 할까?

(1) 협업 필터링 모델과 한계점

다양한 유사도 계산 이해하기

안녕하세요. 에디터 준희, 헤나입니다.

d. <https://deepdaiv.oopy.io/articles/3>

Article

다양한 유사도 계산
이해하기