

2023.03.23 미팅 자료

추천 알고리즘, 제안서

팀원 : 배한성, 김은서, 조현아, 조유진

추천 알고리즘

추천 알고리즘

- 사용자에게 질문을 받고, 질문과 리뷰들간의 코사인 유사도를 구해 유사도가 높은 리뷰가 있는 상품에 **count**를 증가시켜 가장 높은 점수를 가진 상품을 추천하도록 모듈을 구성
- 모든 리뷰에 대해 코사인 유사도를 구하려고 하니 실행 시간이 너무 긴 문제가 있었다.
 - **json**으로 이루어진 리뷰 데이터를 순차적으로 읽었다.
 - **GPU** 환경에서 1 개의 **sentence** 묶음 당 **0.5초** 정도의 시간이 소요되었다.
 - 10 만개의 **sentence**의 문장 임베딩을 구하니 **2시간**의 시간이 소요되었다.
- 따라서 모든 리뷰에 대해 코사인 유사도를 구할 수 없다고 판단하였다.

추천 알고리즘

- 가설 1

- 5점 리뷰만을 사용하여 코사인 유사도를 구한다.
- 하지만, 5점 리뷰만을 사용하여도 실행 시간이 긴 문제가 있었다.

- 가설 2

- 5점 리뷰 중 모든 리뷰를 조사하는 것이 아닌, 몇 개의 리뷰만을 가져오는 방법
- 상품 별 300개의 **sentence**만 가져와 리뷰를 구한다.
- 실행 시간은 3~4분이 소요되었지만, 다양한 사람의 리뷰를 가져올 수 없었다.
- 다양한 사람이 작성한 리뷰를 가져올 수 있는 방법을 생각해볼 필요를 느꼈다.

추천 알고리즘

- 가설 2에 대한 실행 결과 (유사도 조건을 0.675로 하였다.)

```
query: 가벼운 노트북 추천해줘  
run time : 198.11508440971375  
추천하는 상품은 LG전자 그램15 15ZD90N-VX70K 입니다.  
점수는 18 입니다.
```

```
query: 딥러닝에 적합한 노트북 추천해줘  
run time : 199.1461362838745  
추천하는 상품은 삼성전자 오디세이 NT850XBW-XD7A 입니다.  
점수는 3 입니다.
```

추천 알고리즘

- 가설 2에 대한 실행 결과 (유사도 조건을 0.675로 하였다.)

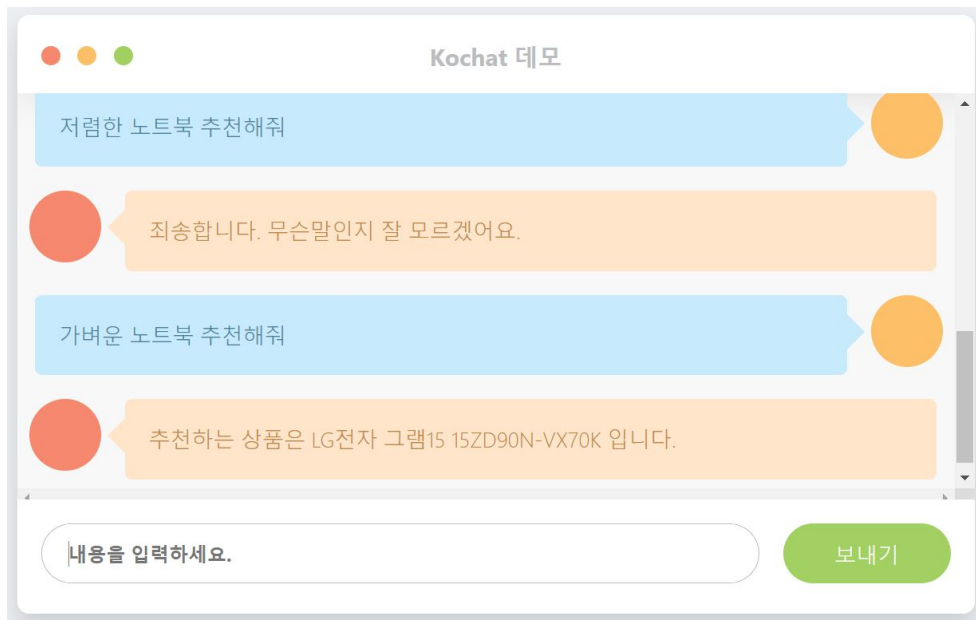
```
query: 삼성 노트북 추천해줘  
run time : 196.25078344345093  
추천하는 상품은 삼성전자 노트북 플러스2 NT550XDA-KC35W 입니다.  
점수는 18 입니다.
```

```
Process finished with exit code 0
```

```
query: 성능 좋은 노트북 추천해줘  
run time : 197.40083837509155  
추천하는 상품은 삼성전자 오디세이 NT850XBW-XD7A 입니다.  
점수는 29 입니다.
```

추천 알고리즘

- 챗봇 API를 통한 답변



추천 알고리즘

- DB를 통해 리뷰를 가져온다고 하여도 약 10만개 가량의 리뷰에 대해 코사인 유사도를 구하는 것은 상당한 시간이 소요될 것이다.
- 리뷰를 읽어오는 시간을 줄이는 방법 찾아볼 예정
 - 다중 Thread 환경에서 Review를 읽어오는 방식을 구현해볼 예정
 - 여러개의 스레드에서 동시에 리뷰를 읽는다면 빠르게 읽을 수 있을 것