
KNOW 직업추천 대회 User-based CF로 분석 결과

김지선

User-based CF 알고리즘으로 분석

- CF 알고리즘을 직업추천 데이터에 적용



CF 적용 프로세스

1. 텍스트 데이터와 숫자 데이터를 분리(학습 데이터)
2. 전체 텍스트 데이터에 대하여 토큰화 해서 단어 모델 생성
3. 텍스트 데이터에 각각에 대해 모델 돌리기
 - 각각의 행, 열 값에 대한 토큰화
 - 토큰화된 값을 모델에 돌리기
4. 임베딩 완료
5. 임베딩한 텍스트 데이터를 숫자 데이터에 붙이기
(=설문 응답 결과를 모두 벡터화)
6. 구직자와 재직자의 설문 응답의 상호 유사도 행렬 생성
7. 구직자와 가장 비슷하게 응답한 재직자의 직업을 추천(구직자에게)

지금 진도 유사도 행렬 이용

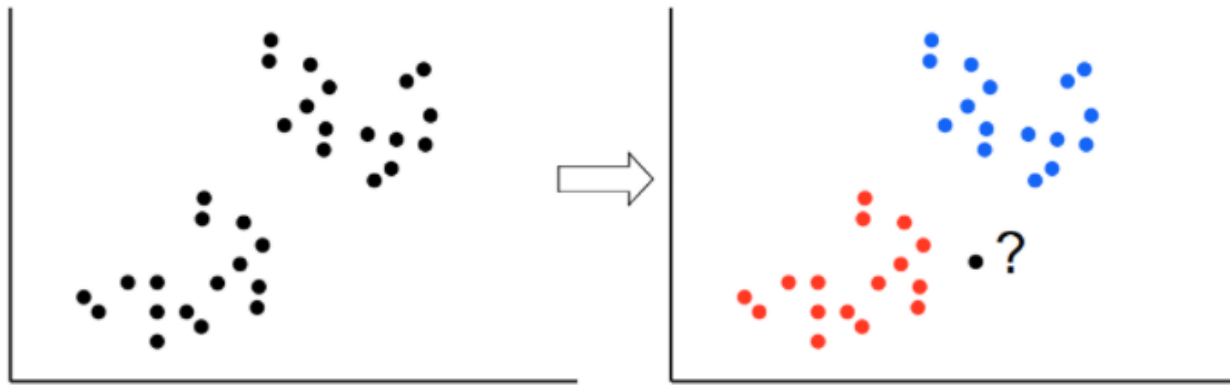
- 유사도 행렬 기반 예측 하면 끝남

	구직자의 응답 (idx=0)	구직자의 응답 (idx=1)	구직자의 응답 (idx=2)
재직자의 응답 (idx=0)	0.32(A)	0.67	0.54
재직자의 응답 (idx=1)	0.55	0.87	0.23
재직자의 응답 (idx=2)	0.74	0.32	0.89(B)

- A-구직자 0과 재직자 0의 설문 응답 유사도
- B-구직자 2에게 재직자 2의 직업 추천 !!
 - ⇒ 구직자 idx=2와 재직자 idx=2의 응답이 가장 비슷하기 때문
- idx=user 식별
- 구직자-test data
- 재직자-train data

KNN으로 직업 추천

- K-NN은 거리기반 분류(유클리디안 거리 기준)
 - k개는 홀수
 - 표준화가 필요
- 결국에 유사한 유저의 직업을 추천하는 아이디어는 동일



적용 결과

- CF로 예측한 결과

2022-03-28 05:32:57	0.0599251768
	0.0630750808

- KNN으로 예측한 결과

2022-03-28 17:21:19	0.0656647121
	0.0693695109

현재상황 & 문제점

- 정확도가 너무 낮음
 - 적절하지 않은 알고리즘 사용?
 - 기존 알고리즘이 아닌 자체 아이디어를 구현해서?
- (1) 라이브러리 핸들링 가능
 - ⇒ 다음 단계 (활용능력이 떨어져서 분발 필요)

앞으로 할 것

- 한전 공모전 참여

- ⇒ 지금하고 있는 것에 집중하는게 나아 보임

- +지금하고 있는 것-머신러닝 스터디, 학부 공부, 교수님이 시키시는 것

- MF로 무비렌즈 데이터 분석