

데이터분석캡스톤디자인

13주차 수행보고

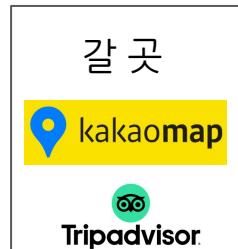
Khupid 조

산업경영공학과 김동혁
관광학과 류연주
산업경영공학과 유정수

플로우

크롤링

유저, 장소, 평점, 리뷰 등



리뷰
영어로 번역
papago

리뷰

평점

리뷰 태깅
형용사 추출

word2vec

클러스터링

kmeans

matrix 생성
(정규화)

User

Place

모델링

CBF
cosine similarity

결과 합산 후 랭킹

Recommend

상위 랭크 장소들을
해당 User에게 추천

rating data

NCF

neural collaborative filtering

place matrix

	0	1	2	3	4	5	6	7	8	9	...	110	111	112	113	114
p_id																
721	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.0
1832	0.053308	0.000000	0.000000	0.000000	0.0	0.000000	0.046086	0.0	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.0
741	0.034111	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.055405	...	0.000000	0.078135	0.000000	0.000000	0.0
218	0.042261	0.000000	0.000000	0.000000	0.0	0.077576	0.182678	0.0	0.079761	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.0
3276	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.0
...
3793	0.003882	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.018916	...	0.179379	0.026677	0.000000	0.000000	0.0
2809	0.016743	0.000000	0.000000	0.000000	0.0	0.006147	0.043423	0.0	0.018959	0.024475	...	0.000000	0.080536	0.000000	0.010435	0.0
3297	0.016136	0.002938	0.001383	0.00235	0.0	0.006439	0.021229	0.0	0.007945	0.010256	...	0.003242	0.043390	0.002137	0.006559	0.0
445	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.039573	0.0	0.000000	0.000000	...	0.000000	0.000000	0.055763	0.000000	0.0
3592	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.060478	0.0	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.0

user matrix

	0	1	2	3	4	5	6	7	8	9	...	110	111	112	113	114
u_id																
5351	0.025431	0.0	0.0	0.000000	0.027046	0.000000	0.036643	0.0	0.031998	0.020653	...	0.000000	0.058252	0.000000	0.000000	0.0
3024	0.017392	0.0	0.0	0.000000	0.004624	0.00266	0.021927	0.0	0.005471	0.007062	...	0.006697	0.004980	0.026483	0.000000	0.0
5941	0.015398	0.0	0.0	0.003558	0.010164	0.000000	0.048199	0.0	0.008017	0.005175	...	0.004907	0.014595	0.000000	0.001655	0.0
13580	0.019210	0.0	0.0	0.000000	0.000000	0.000000	0.083038	0.0	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.059865	0.0
6953	0.000000	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.0
...
18726	0.000000	0.0	0.0	0.000000	0.000000	0.000000	0.114951	0.0	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.0
10858	0.078430	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.0
10956	0.055347	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.0
21694	0.022082	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	...	0.000000	0.151741	0.000000	0.000000	0.0
18178	0.000000	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.0

맛집 (음식점, 카페)

```
result.sort_values('result', ascending = False).head(10)
```

user 28에 대한
최종 추천 결과



p_id	name	address	cbf_score	ncf_scores	result
4142	중앙해장	서울특별시 강남구 영동대로86길 17 육인빌딩	0.880745	0.604347	0.532276
1068	동경산책	서울특별시 성북구 보문로34길 45	0.915014	0.552113	0.505191
1712	뽕탄	서울특별시 용산구 백범로99길 50	0.919703	0.547507	0.503543
3468	온달 왕 돈까스	서울시 성북구 동선동1가 2-9	0.697894	0.704212	0.491465
1658	모리돈부리	서울특별시 관악구 관악로12길 6	0.876423	0.522881	0.458265
2704	웨이크썬 두타점	서울특별시 중구 을지로6가 18-12	0.892957	0.509402	0.454874
4412	카레시오	서울특별시 관악구 관악로14길 28	0.799053	0.568145	0.453978
1587	메이크샐러드	서울특별시 동대문구 휘경로2길 16 1F	0.805195	0.556817	0.448346
747	뉴질랜드스토리	서울특별시 송파구 송파동 32-1	0.691385	0.628268	0.434375
4772	토끼정	서울특별시 강서구 하늘길 38 롯데몰 김포공항점 MF	0.875528	0.495252	0.433607

갈 곳 (관광 명소, 공원, 불링장, 노래방, 만화카페, 방탈출 카페, 멀티방)

Contents-based - 단순 행렬곱 <user_id = '진* 이'>

1	2	3	4	5	6	7	8	9	...	93	94	95	96	97	98	99	p_id	total_count	preference
0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.037037	...	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.0	f239	3	0.004115
0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000	...	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.0	b209	0	0.000000
0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000	...	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.0	e810	3	0.002939
0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000025	...	0.000051	0.0	0.0	0.0	0.000000	0.000051	0.0	a211	27	0.000022
0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000	...	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.0	b44	6	0.000257
...
0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000015	...	0.000007	0.0	0.0	0.0	0.000000	0.000005	0.0	1484	41	0.000004
0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000	...	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.0	1485	0	0.000000
0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000296	0.000000	...	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.0	1500	15	0.000091
0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000	...	0.000025	0.0	0.0	0.0	0.000051	0.000017	0.0	1503	27	0.000018
0.0	0.0	0.0	0.0	0.0	0.001372	0.0	0.000000	0.000686	...	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.0	1510	9	0.000256

NCF 모델 결과

0	1
0	8.0 0.420125
1	68.0 0.004540
2	16.0 0.294885
3	27.0 0.106112
4	69.0 0.362821
...	...
1423	804.0 0.000410
1424	1322.0 0.000448
1425	933.0 0.000724
1426	722.0 0.000414
1427	1168.0 0.000389

=

최종 결과

0.0017839710676999104,
0.0,
0.00048695306053530576,
2.146124576770359e-06,
0.00011768787273204276,
3.329086880004263e-05,
0.00015081535900896183,
0.00010731165891724683,
0.0,
0.010317383678096115,
0.0,
0.0004716094132441623,
0.00012700195620014412,
0.0,
0.0002244194127192164,
1.2240620203866347e-05,
0.0,
9.60483081949913e-05,
2.558109955118021e-05,

볼 곳 (영화)

NCF 결과

	0	1
0	31.0	0.008509
1	2.0	0.020827
2	118.0	0.157921
3	68.0	0.021750
4	15.0	0.082053
...
503	431.0	0.009089
504	445.0	0.008397
505	476.0	0.009716
506	490.0	0.009067
507	471.0	0.008932

X

contents-based 결과

	0	1	2	3	4	5	6	7	8	9	...	91	92	93	94	95	96	97	98	99	preference
0	0.0	0.0	0.004975	0.0	0.02439	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0	0.02439	0.000059
1	0.0	0.0	0.004975	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0	0.02564	0.000044
2	0.0	0.0	0.009524	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.000117
3	0.0	0.0	0.006160	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.095238	0.047619	0.0	0.0	0.0	0.0	0.0	0.047619	0.000100
4	0.0	0.0	0.014340	0.0	0.000000	0.0	0.0	0.0	0.058824	0.0	...	0.0	0.000000	0.058824	0.0	0.0	0.0	0.0	0.0	0.000000	0.000141
...
503	0.0	1.0	0.154229	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	...	0.0	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
504	0.0	0.0	0.199005	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
505	0.0	0.0	0.154229	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
506	0.0	0.0	0.084577	0.0	1.000000	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
507	0.0	0.0	0.194030	0.0	0.000000	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

```
user_1_ncf.sort_values('result', ascending=False)
```

user1에 대한
최종 추천 결과



	0	1	result
154	296.0	0.058601	2.224491e-04
115	1.0	0.077351	2.181064e-04
105	30.0	0.044600	2.122365e-04
96	181.0	0.060587	1.948521e-04
158	169.0	0.048466	1.890429e-04
...
41	49.0	0.008757	6.075337e-07
0	31.0	0.008509	4.988874e-07
81	368.0	0.015640	3.871300e-07
75	97.0	0.009120	2.257463e-07
507	471.0	0.008932	0.000000e+00

다음주 할 일

동혁

신규 사용자 이용시
인터페이스 및 로직 구현

NCF 모델 성능 높이기

연주

중복된 데이터 체크 (kakao
placs)

행렬곱 -> 코사인 유사도
변경

NCF 모델 성능 높이기

정수

클러스터링 &NCF 성능
높이기

공통

- 보고서 앞 부분 작성 (~데이터 전처리)
- 데이터 취합 (word cluster, user/place matrix)
- Test 준비