

ZERO-BASE DS School

DL Project



위치와 사용자 취향을 반영한 데이트 경로 추천 시스템

2조 이동희, 이슬기, 장문정



목차

1. 주제
2. 배경/기획의도
3. 사전 자료
4. 구현 모델
5. 향후 진행 방안



주제

사용자 위치 기반

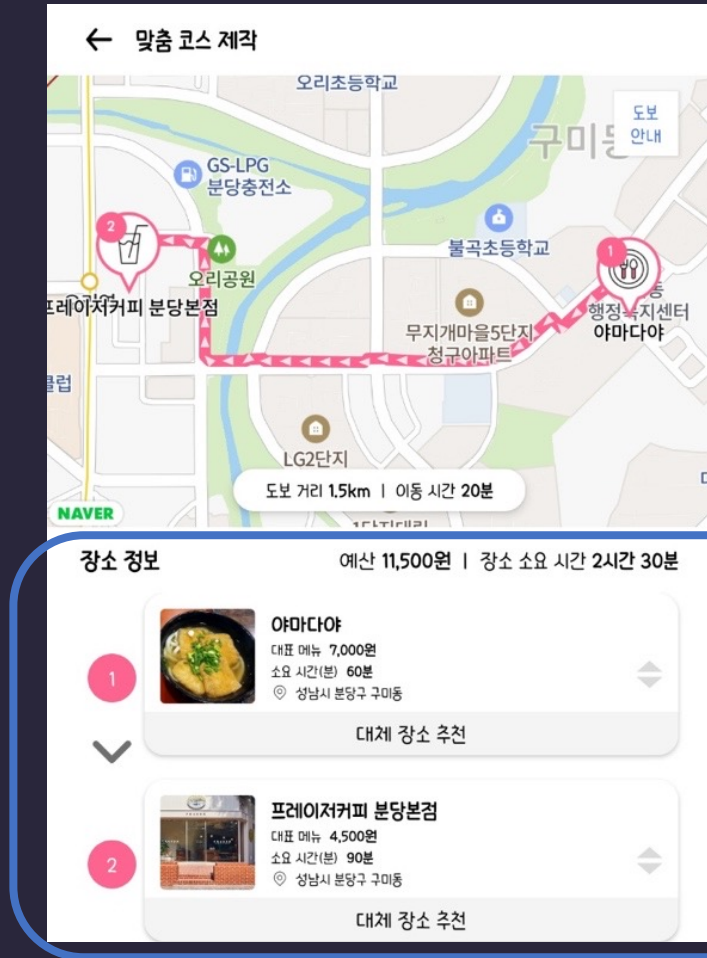
사용자 취향을 반영

주변의 명소, 맛집, 카페를 잇는 데이트 코스 추천 시스템 개발



사전자료

모바일앱 '완벽한 하루'



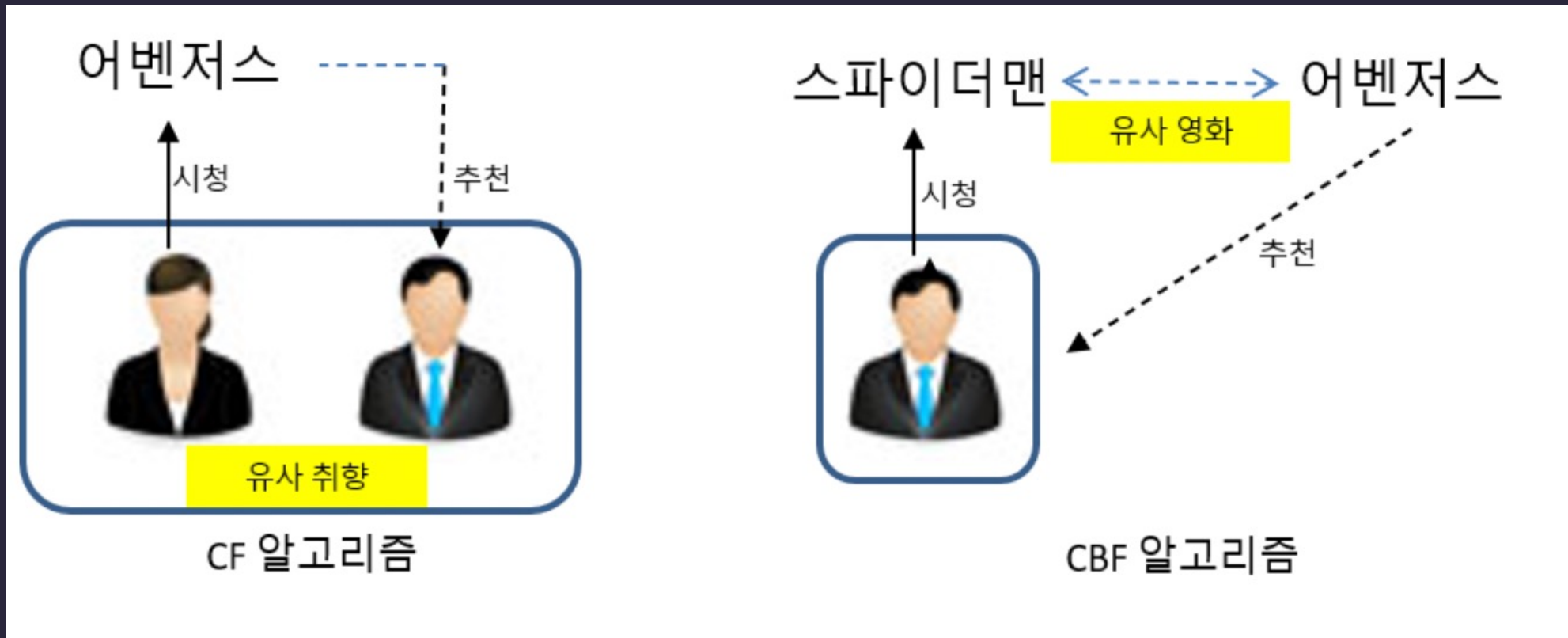
Place 추천

-> HOW??



구현모델

Contents-based Filtering vs Collaborative Filtering

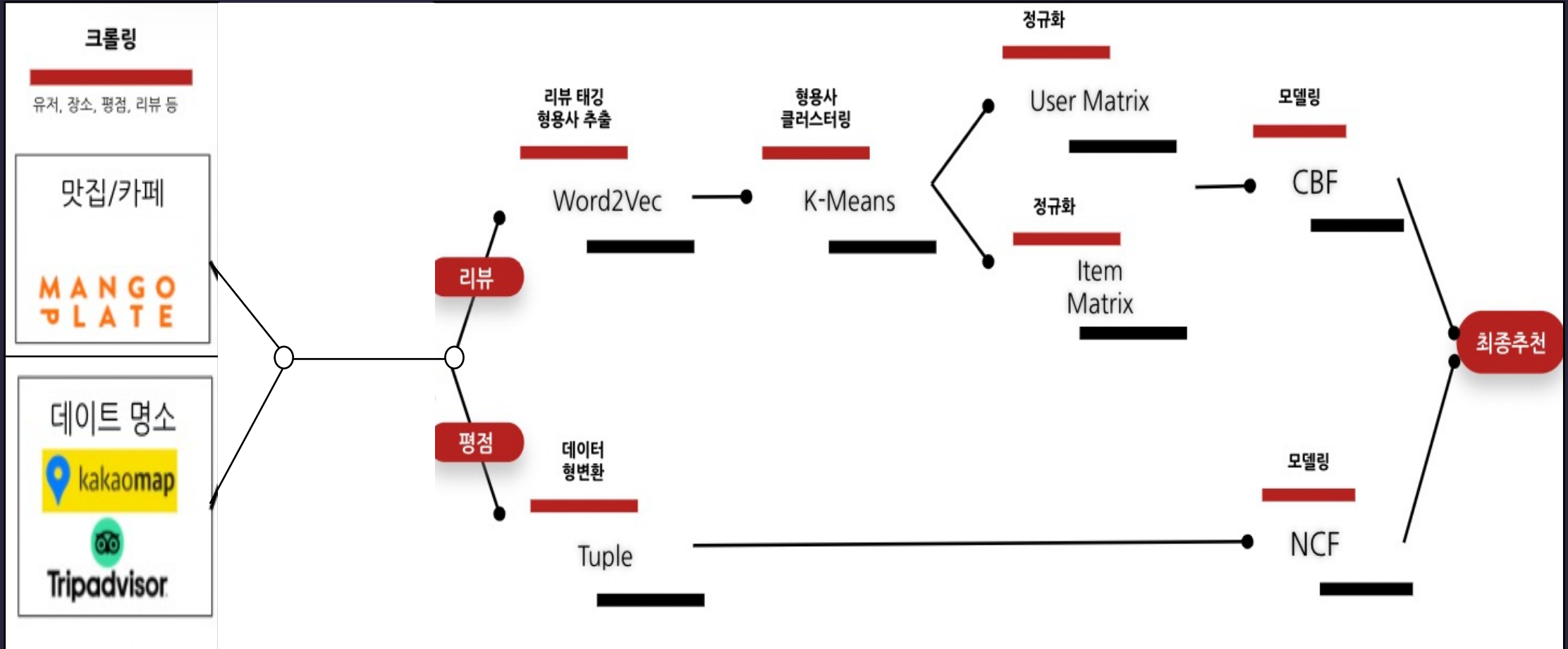




Hybrid model
= Contents-based Filtering (CBF; user, item based)
+
Neural Collaborative Filtering (NCF)



추천 알고리즘 상세 라인



출처 : <https://github.com/DAC-KHUPID/seoul-date-course-recommendation>



구현모델

Konlpy 를 통해 가게별 tag 추출

eat_info									
	name	address	category	price	rvw_cnt	rating	p_id	reviews	tags
0	원조할아버지순두부	서울특별시 광진구 자양로 324	한정식 / 백반 / 정통 한식	5000	72	4.5	553	예전 초당할머니 순두부는 살짝 변해서여길 못 따라옴원순두부는 맛과 향이 살짝 비릴수 ...	초당,순두부,순두부,향,이집,전혀,아차산,등산,필수,코스,부집,늘,회전,십분,이상,...
1	서북면옥	서울특별시 광진구 자양로 199-1	국수 / 면 요리	5000	94	4.2	383	서울특별시 광진구 구의2동에 위치한 서북면옥에서 물냉면(9천원)과 접시만두 6조각(...	서울특별시,광진구,구의,동,서북,면옥,물냉면,과,접시,만두,물냉면,육수,습습,소고기,...
2	케도에오르다	서울특별시 광진구 광나루로 416 건희빌딩 1F	기타 한식	5000	40	4.0	78	기본에 충실해서 더 맛있었던 핫도그였습니다. 들어가서 음식이 나오기까지는 꽤 시간이 ...	핫도그,문,준비,고려,핫도그,특성,아쉬움,소세지,굽기,적,당선,양은,간식,광진구,투...
3	콩이네	서울특별시 광진구 긴고랑로9길 28 중국제일시장 103호	기타 한식	5000	13	4.0	715	이사온 첫날 점심은 짜장면인데 집에서 먹기 싫어서밖으로 나와 망플을 컵니다..1위기 ...	은,첫날,짜장면,망플,장안,분식집,사장,대한민국,요식업,친절,장안,떡볶이,튀김,오뎅...
4	건대우동집	서울특별시 광진구 아차산로 201 동성빌딩 1F	기타 한식	5000	9	4.0	45	혼밥하기 좋은곳.김밥도 괜찮고 잔치국수도 맛 음하니 괜찮다.. 잔치국수와 건대김밥 주 ...	혼밥,김밥,잔치국수,잔치국수,김밥,잔치국수,다대기,음,김밥,계란,당근,단무지,위주,...



구현모델

Konlpy 를 통해 user별 tag 추출

```
[ ] eat_review_by_user
```

	reviews	tags
715	예전 초당할머니 순두부는 살짝 변해서여길 못따라옴원순두부는 맛과 향이 살짝 비릴수 ...	초당,순두부,순두부,향,이집,전혀,동치미,향,맛뒤,육수,혼밥,김밥,잔치국수,쫄깃함신...
453	아차산 등산 후 내려오는 길에 필수코스처럼 되어버린 두부집. 늘 대기가 조금씩 있는...	아차산,등산,필수,코스,부집,늘,회전,십분,이상,적,시,두부,물이,두부,순두부,콩국...
1080	<다른메뉴 없고 순두부만 파는 집>이 집의 매력포인트 중 하나는 막걸리가 정말 싸다...	두부,집의,매력,포인트,막걸리,막걸리,반찬,두부,합리,갑,스시,마카,만철천원,입성,...
1479	등산 후 맛보는 기가막힌 #원조할아버지순두부- 직접 만든 유기농 100% 순두부- ...	등산,기,유기농,두부,순환,아차산,등산,서울,시내,풍경,눈,하산,할아버지,부집,두부...
954	드디어 먹어본 아차산 순두부!! 같이 가 준 친구들이 이미 전에도 맛본 경험이 있어...	아차산,순두부,준,전,경험,등산,해도,두부,부가,일인,순두부,막걸리,두부,두부,백설...
...
537	오픈매장인거 보고 왔다갔어요!개인적으로 쉬림프파스타 정말 맛있었어요!!또 들를래요 ...	쉬,림프,파스타
1597	오픈한지 별로 안된 가게라서 가게에 사람들이 많이 없더라고요...사실 들어가기 두려...	한지,남자,돌이서,로마,휴일,흑백영화,빔,프로젝트,노래,술,취하,주시,레스토랑,검,...
603	자양동에 이런 분위기 와인펍이 들어올줄은 몰랐네요 엄청 고급집니다. 소개해보면 국내...	자양동,와인펍,고급,소개,국내,유명,셰프,개발,참여,와인,고급,레스토랑,비주
655	소리 너무너무너무 시끄러워요. 클럽마냥 소리질러야 대화가 됩니다~ 야경은 예뻐요.	소리,클럽,마냥,소리,대화,야경
58	비비드아워로 방문.종류는 아주 다양하다거나 신라 정도의 퀄리티는 아니지만 뷰가 예쁘...	비비드,워,신라,뷰,언어,등,보더,워커힐,뷔페,조식,차라리,디저트,낮다,중점,술,마시기

1730 rows x 2 columns



구현모델

한국어의 미리 학습된 Word2Vec

```
import gensim  
w2v = gensim.models.Word2Vec.load('/content/drive/MyDrive/Colab Notebooks/ko/ko.bin')
```



구현모델

K-means를 통해 tag 클러스터링
200개 group

group_44	group_74	group_36	group_28	group_51	group_52	group_53
의심	가재	버섯	대학생	한강	경상도	중화
소문	호밀	땅콩	중국인	산악	농	등
모독	백합	쇠고기	극단	동부	서북	하노이
비유	소고기	고추	유목민	중부	강원도	베이징
대접	미역	케이크	대거	어항	사천	마카오
죄송	좁	김치	젊은이	인구	원주	상하이
걱정	귀리	재배	주민	본토	청주	중국
고집	술	돼지고기	명의	지진	흥	베트남
항의	허브	맥주	일본인	국립공원	무안	위안
극찬	제과	식용	경력	이남	평안	광저우
친절	고사리	생선	종업원	겨울철	광주	콩
려	딸기	빵	어린이	동북	읍	쓰촨
착각	약용	커피	후배	근해	대동	타이
찬양	쌈	우유	관광객	어선	세종	징
배려	파슬리	감자	여성	홍수	조선	0
하자	메밀	소금	신분	강풍	지리산	0
소식	청어	초콜릿	미국인	수로	평안도	0
감동	제조법	열매	신입생	남부	이천	0
원한	부추	가루	군인	지역	경주	0
		버터	명과	경계	전라도	0



구현모델

장소 matrix 생성

```
[ ] res_matrix.head()
```

	group_29	group_0	group_1	group_2	group_3	group_4	group_5	group_6	group_7	group_8	group_9	group_10
p_id												
553	0	0	0	0	0	0	0	0	0	0	0	0
383	0	0	0	0	0	0	0	0	0	0	0	0
78	0	0	0	0	0	0	0	0	0	0	0	0
715	0	0	0	0	0	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0	0	0	0

user matrix 생성

```
[ ] user_matrix.head()
```

	group_29	group_0	group_1	group_2	group_3	group_4	group_5	group_6	group_7	group_8	group_9	group_10	g
715	1	0	0	0	0	0	0	0	0	0	1	0	
453	8	2	1	0	2	1	1	0	0	0	1	2	
1080	6	0	0	2	5	1	0	0	0	0	0	0	
1479	3	0	0	1	1	1	1	0	0	0	0	0	
954	18	2	0	0	11	1	11	0	1	0	5	2	



Neural Collaborative Filtering (NCF)

- 2017년 공개된 추천 알고리즘
- Collaborative Filtering에 Neural Network를 접목
- Matrix Factorization의 한계를 보완하기 위해 연구 됨



구현모델

NCF 모델 성능

...							
	F1@10	NDCG@10	Precision@10	Recall@10	Train (s)	Test (s)	
-----	-----	-----	-----	-----	-----	-----	-----
NeuMF	0.2335	0.7053	0.1578	0.8551	204.7705	2.6812	



향후 진행 방향

보완 및 수정

- CBF 알고리즘 완성 (현재 matrix 데이터 정리 마친 상태)
- NCF (광진구 한개의 구만 적용해 본 상태로 전체 데이터 진행 시 성능개선)
- 지도 표현을 위해 Folium 적용 중
- 웹페이지 추천

감사합니다