

요기요 추천시스템

1조 요기가
배성수 | 이기연 | 구본근 | 양진이 | 이현서

프로젝트 목표

'요기요' 플랫폼에 등록되어 있는 **'마포구'** 지역의
음식점 데이터와 사용자 평가를 수집한다.

수집한 데이터를 통해 사용자와 취향이
유사한 다른 사용자를 찾고,
사용자가 주문할 수 있는 지역 내 음식점을
추천하는 모델을 선정한다.

수집 데이터 전처리

수집과정

- 데이터 수집을 위해 '요기요' 웹사이트 대상으로 크롤링을 진행하였다.
- 배성수 구본근 학생이 예정대로 데이터 크롤링 모델을 생성하였다.
- 데이터 수집은 음식의 카테고리를 기준으로 진행하였다.
- 양진이 이현서 학생이 수집한 데이터를 다시 취합하여 하나의 데이터 Set으로 형성하였다.

이슈사항

- 데이터를 수집하는 과정에서 식당과 고객의 정보가 많은 관계로 수집 시간이 오래걸렸다.
 - >> 팀 구성원 모두가 각자 다른 카테고리를 맡아 수집을 진행하였다.

추천 시스템에 적용할 모델과 목표

적용모델

- 코사인 유사도
- SVD & NMF
- Neural Network

목표

- 해당 프로젝트에 가장 적합한 모델이 어떤 것 인지 파악한다.
(기준: 위 세 모델을 중 가장 정확도가 높은 모델 = 가장 적합한 모델)

'코사인 유사도'를 통한 추천시스템

구현과정

- 수집한 데이터를 `nn.CosineSimilarity()`를 사용하기 적합한 형태로 가공한다.
- 사용자마다 평점 추정값을 구하는 함수를 생성하여 평점 값을 예상한다.
- 사용자의 평점 추정값이 높게 나온 식당 중 방문한 적이 없는 가게를 반환한다.
- 그 수는 3곳으로 제한한다.

정확도 측정(Mean Squared Error)

MSE: 9.966770847183565

이슈사항

- 모든 가게에 평점을 준 유저가 존재하지 않아 결측치를 제외하고 진행할 수 없었다.
- 결측치가 많아 결측치를 처리해야 코사인 유사도 진행이 가능했다.
- 결측치를 유저가 준 모든 가게의 평점의 평균으로 넣었을 때, mse값은 13으로 높은 값이 나왔다.
- 결측치를 0으로 설정하여 mse값을 10으로 개선하였다.
- 평가가 10개 이상인 유저와 식당으로 데이터를 축소하여 mse값을 9점대로 개선하였다.

'SVD & NMF'를 통한 추천시스템

구현과정

- 수집한 데이터를 Array행렬의 형식으로 가공한다.
- SVD, NMF 모델학습을 각각 진행한다.
- Loss 계산함수를 생성하여 각 모델의 SSE, RMSE, MSE 값을 구한다.

정확도 측정(SSE, RMSE, MSE)

```
SVD Error(ignoreing zero values): SSE = 121049.20341906235, RMSE = 1.3421553726608433, MSE = 1.801381044362367
NMF Error(ignoreing zero values): SSE = 267900.5549872832, RMSE = 1.9966807592591547, MSE = 3.9867340543957144
```

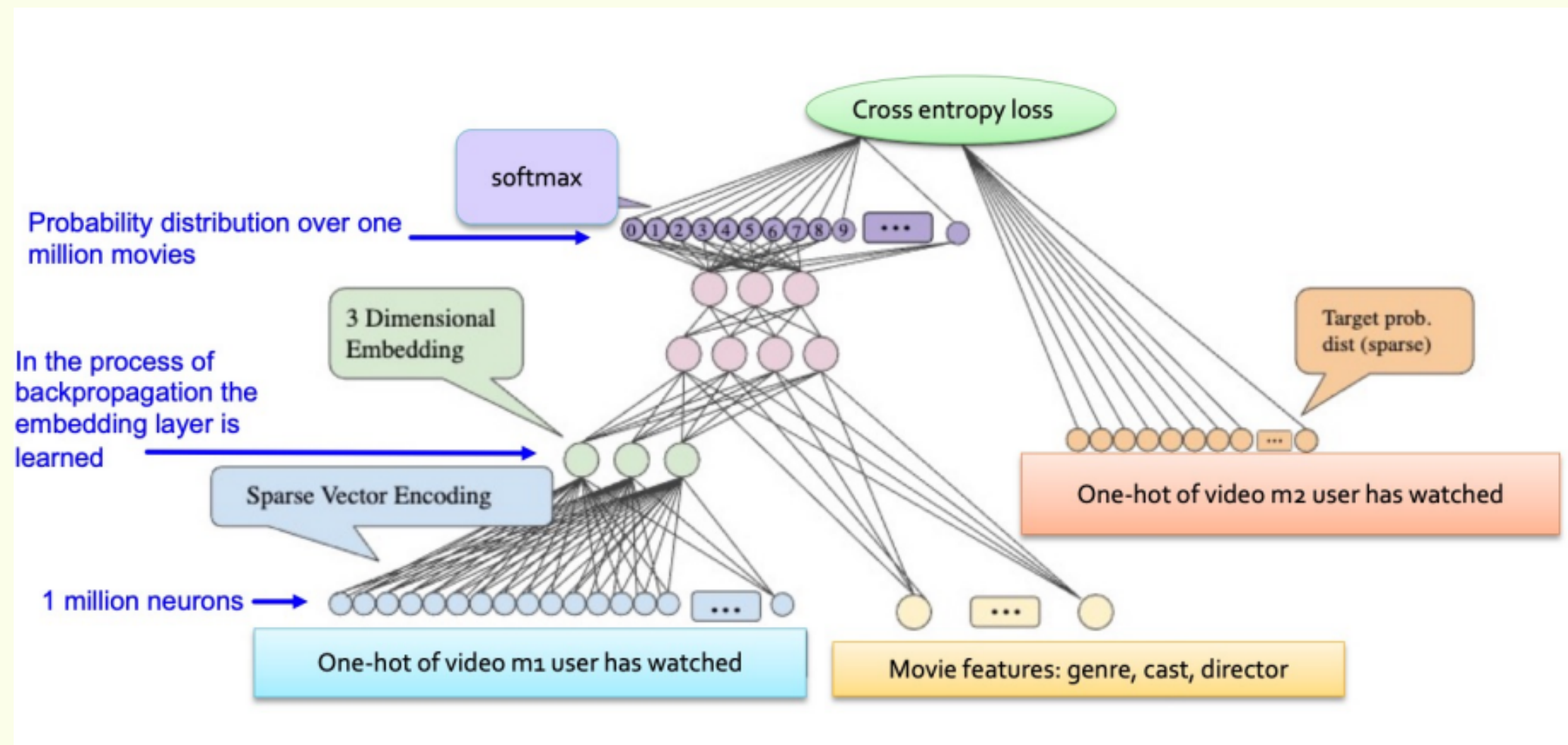
비교결과

- SVD 모델의 성능이 더 좋다는 것을 위 결과를 통해 알 수 있다.

'인공 신경망'을 통한 추천시스템

구현과정

- 수집한 데이터를 Index작업을 한 후 Pairs로 형성한다.
- Train set : Test set = 9 : 1 로 임의 설정한다.
- 모델을 생성한다.



	layer	batch	log interval	learning rate	epochs	last loss	Training set	Test set
1	50	256	3400	0.01	20	6.074473	43657/8741950(0%)	625/108190(1%)
2	50	1024	850	0.005	20	-	48943/8741950(1%)	533/108190(0%)
3	50	2048	425	0.001	20	6.018146	48936/8741950(1%)	560/108190(1%)
4	50	2048	425	0.0015	20	6.016363	49151/8741950(1%)	569/108190(1%)
5	50	2048	425	0.002	20	6.019795	49165/8741950(1%)	570/108190(1%)
6	50	2048	425	0.0025	20	6.018225	49041/8741950(1%)	569/108190(1%)
7	50	2048	425	0.003	20	6.017991	49015/8741950(1%)	551/108190(1%)
8	50	2048	425	0.0035	20	6.019066	49122/8741950(1%)	550/108190(1%)
9	50	2048	425	0.004	20	6.019514	49445/8741950(1%)	555/108190(1%)
10	50	2048	425	0.0045	20	6.02088	49407/8741950(1%)	562/108190(1%)
11	50	2048	425	0.005	20	6.02097	49367/8741950(1%)	548/108190(1%)
12	50	2048	425	0.1	20	6.220864	36567/8741950(0%)	423/108190(0%)
13	50	2048	425	0.01	20	6.022983	48586/8741950(1%)	522/108190(0%)
14	50	2048	425	0.007	20	6.020408	49181/8741950(1%)	572/108190(1%)
15	50	2048	425	0.006	20	6.024334	49345/8741950(1%)	560/108190(1%)
16	5	2048	425	0.007	20	6.028742	48457/8741950(1%)	518/108190(0%)
17	5	2048	425	0.1	20	6.193419	37520/8741950(0%)	450/108190(0%)
18	5	2048	425	0.01	20	6.054567	47074/8741950(1%)	522/108190(0%)
19	5	2048	425	0.001	20	6.035012	48633/8741950(1%)	541/108190(1%)
20	5	2048	425	0.003	20	6.030637	48440/8741950(1%)	542/108190(1%)
21	5	2048	425	0.007	40	6.047465	47953/8741950(1%)	569/108190(1%)
22	5	2048	425	0.003	40	6.04223	48005/8741950(1%)	540/108190(0%)
23	50	2048	425	0.003	40	6.019107	49583/8741950(1%)	551/108190(1%)
24	50	2048	425	0.007	40	6.017955	49359/8741950(1%)	546/108190(1%)
25	50	2048	425	0.008	40	6.021475	49073/8741950(1%)	559/108190(1%)
26	50	2048	425	0.003	80	6.016671	49724/8741950(1%)	557/108190(1%)
27	50	2048	425	0.007	80	6.022097	48684/8741950(1%)	548/108190(1%)
28	50	2048	425	0.001	80	6.01793	49501/8741950(1%)	553/108190(1%)
29	5	2048	425	0.003	80	6.038917	48383/8741950(1%)	520/108190(0%)
30	5	2048	425	0.007	80	6.048497	47642/8741950(1%)	537/108190(0%)

각 모델 추천 결과(한명의 사용자 대상)

SVD

	Rating	prediction	Restaurant
603	0.0	2.263789	꼬치꼬치양꼬치
2	0.0	2.176248	미스터피자-상암점
257	0.0	2.144737	명태요리전문점김명태-서교점
444	0.0	2.032146	신촌돈부리모노-본점
464	0.0	1.985538	푸라닭-망원점
590	0.0	1.786323	식공간-연남점
394	0.0	1.654453	이불덮밥-상암DMC점
470	0.0	1.593873	로타리치킨홍대상회
282	0.0	1.412486	상수,곰도리탕
81	0.0	1.407305	쿠벤일식도시락

NMF

	Rating	prediction	Restaurant
140	0.0	3.340058	연희옥-연희본점
257	0.0	3.283451	명태요리전문점김명태-서교점
2	0.0	2.997640	미스터피자-상암점
238	0.0	2.798416	신촌해물떡찜
504	0.0	2.759761	처갓집양념치킨-홍대점
631	0.0	2.758999	호세야오리바베큐-합정점
40	0.0	2.742423	식스티즈버거
237	0.0	2.725164	삼겹편치-마포본점
83	0.0	2.645004	USA존슨부대찌개-수색점
329	0.0	2.345339	동대문엽기떡볶이-신촌점

COS SIM

	predict rate
Restaurant	
꽃게나라 간장게장	0.306709
두남자의볶음밥 & 포슈아베트남쌀국수	0.304825
대죽장	0.292403
치킨플러스-성산점	0.285083
족발야시장-신촌점	0.284898
또래오래-마포망원점	0.278045
스시앤-마포점	0.275958
벚꽃스시-본점	0.274341
KFC-홍익대점	0.273157
안동장	0.271806

NEURAL NETWORK

	Mean_Rating	similarity
Restaurant		
피자파는집-홍대점	1.074402	0.875406
처갓집양념치킨-신촌점	1.470360	0.871965
김밥천국-망원점	1.335886	0.865848
한끼한상-홍대점	0.297628	0.861926
손오공마라탕-서교점	1.366200	0.856168
신미경홍대닭갈비3	0.853650	0.843614
굽네치킨&피자-북가좌1호점	1.704254	0.837528
처갓집양념치킨-서교점	1.105720	0.829281
7번가피자-서교망원점	1.080657	0.823512
청년치킨-서교점	1.457313	0.820213

각 모델 추천 결과(한명의 사용자 대상)

NEURAL NETWORK with Feature

사용자가 방문하여 매긴 식당의 평점

In [58]:	user_restaurant.iloc[3][user_restaurant.iloc[3].notna()]		
Out [58]:	Restaurant		
	굽네치킨&피자-북가좌1호점	5.0	
	굽네치킨&피자-연남점	4.0	
	롯데리아-망원점	5.0	
	매운국물떡볶이밀방떡-홍대입구점	5.0	
	불타는형제들연탄불고기	5.0	
	섬이자카야	5.0	
	신미불닭발-서울본점	5.0	
	야한곱창-홍대점	5.0	
	육이네김치찌개와직화삼겹-신촌점	5.0	
	육회바른연어-홍대점	5.0	
	이가네간장&양념게장	4.0	
	정성이가득찬집밥-이대점	5.0	
	키사-디저트&눈꽃빙수	5.0	
	타이반점	5.0	
	피자파는집-홍대점	5.0	
	행복왕갈비	5.0	
	홍대미남보쌈족발-서교점	5.0	
	홍리마라탕	5.0	
	후라이드참잘하는집-마포점	5.0	
	Name: _**님, dtype: float64		

사용자에게 추천된 식당 리스트

_**님		
	Mean_Rating	similarity
Restaurant		
매일삶는족발-서교점	4.801587	0.447430
푸라닭-연희점	4.844615	0.446527
굽네치킨&피자-연남점	4.848000	0.443834
김기엽초밥좋은날-홍대점	4.835585	0.439301
곱깨비직화불닭치킨-홍대입구점	5.000000	0.438715
홍푸드1인분&갈비찜	4.547330	0.438217
육회바른연어-홍대점	4.898773	0.437487
곱깨비직화삼겹-가좌점	5.000000	0.436011
신릉푸마라탕-상암점	4.707814	0.435761
신미경홍대닭갈비3	4.707682	0.434518

각 모델 정확도 비교

각 모델 정확도

- 코사인 유사도: 9.96...(MSE)
- SVD: 1.80...(MSE)
- Neural Network: 6.01...(Cross Entropy)

고려사항

- MSE는 편차를 지나치게 확대, 오차가 가장 적음.
- 모든 모델에서 평점이 0점인 것을 제외하였다.

최종 적합 모델 판정

적합모델

- 인공지능망 모델이 가장 적합하다고 판단하였다.

이유

- 인공지능망 모델에서 기존 사용자가 좋게 평가한 식당과 추천받은 식당의 유사도가 가장 높았다.
- 편차에 대한 수치적인 정보만으로 적합성을 판단하기에는 기준이 모호하였고 실제로 값의 영향력이 크지않았다.

결론

- 데이터의 유형, 데이터 수, 얻고자하는 값에 따라 적합모델이 다르다는 것을 경험할 수 있었다.