

데이터분석캡스톤디자인

9주차 수행보고

Khupid 조

산업경영공학과 김동혁
관광학과 류연주
산업경영공학과 유정수

한글 형용사 모델 클러스터링 및 matrix 생성

movie_id	snt_review	tags
0	뜨지 못한 명작 사회라는 거대한 틀에서 표류하고 있는 두 사람의 극복기. 정말관심은 ...	['거대하다', '괜찮다', '신선하다', '이렇다', '좋다', '따뜻하다', '...
1	이 영화를 제대로 이해할 수 있는 사람이 몇이나 될까. 아주 고유한 개인들의 기가막...	['비슷하다', '힘들다', '힘들다', '적절하다', '적절하다', '재미있다', '...
2	영화의 사회적 소명을 다하는 작품\n부끄럽지만 처음으로 그리스의 현 상황에 대해 진...	['부끄럽다', '신선하다', '어떨다', '안되다', '여전하다', '무겁다', '...
3	잘 만들어진 스페인 스릴러 영화가 여기 또 있네.\n(스페인 영화는 볼 때마다 스...	['당하다', '신선하다', '매끄럽다', '재있다', '이쁘다', '색시하다', '...
4	왜 우리는 재난 영화를 이렇게 담백하고 몰입감 넘치게 만들지 못 하는가 담백한 연출...	['담백하다', '넘치다', '담백하다', '안타깝다', '끝내주다', '엄청나다', '...
...
665	'조니엡스럽다'는 형용사도 가능할 듯! 디즈니월드 놀이기구를 위한 영화가 이렇게 ...	['스럽다', '가능하다', '괜찮다', '좋다', '훌륭하다', '즐겁다', '넘...

0번 영화에서 추출된 형용사 중 3번 클러스터에 해당되는 형용사가 8개라는 의미

=> 협업 필터링 모델에 사용될
아이템 형용사 매트릭스 생성
(계속 튜닝 진행 예정)

cluster 40
['진중하다', '묵직하다', '식상하다', '그림자', '미치다', '치열하다']

cluster 41
['허전하다', '허탈하다', '다름있다', '친하다', '불필요하다', '찰지다', '상쾌하다', '느슨하다', '한결같다']

cluster 42
['바람직하다', '유연하다', '실었다', '웅하다', '무사하다', '현명하다', '틀림없다']

cluster 43
['무능하다', '엷었다', '억다', '침착하다', '간결하다', '따스하다', '발칙하다']

cluster 44
['담담하다', '정갈하다', '맑다', '불완전하다', '흔흔하다', '시급하다']

cluster 45
['쉽다', '조마조마하다', '야무지다', '유희하다', '어둡다', '광활하다', '놀래다', '악이다', '너무하다']

cluster 46
['명청하다', '듬직하다', '기막히다', '부자연스럽다', '아기자기하다', '늙다', '무력하다']

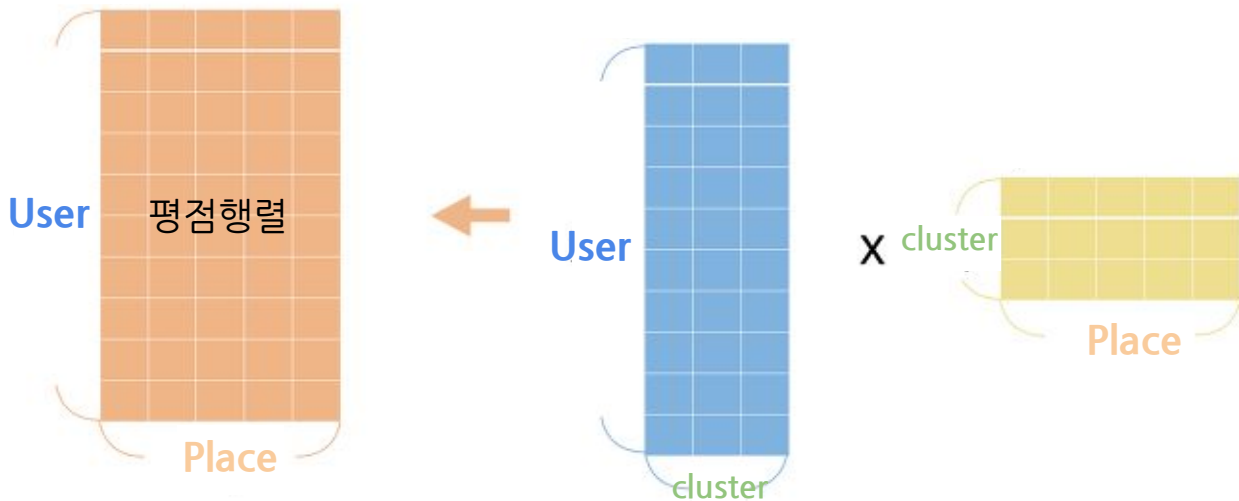
cluster 47
['곰직하다', '서운하다', '빈약하다', '싱겁다', '민망하다', '들뜨다', '장황하다', '무리다', '연하다', '아슬아슬하다']

cluster	0	1	2	3	4	5	6	7	8	9	...	40	41	42	43	44	45	46	47	48	49
0	0.0	9.0	1.0	8.0	0.0	0.0	1.0	5.0	2.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0
1	0.0	6.0	0.0	1.0	2.0	0.0	4.0	4.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0
2	0.0	1.0	0.0	1.0	0.0	0.0	2.0	5.0	3.0	0.0	...	9.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	9.0	0.0
3	0.0	2.0	0.0	0.0	1.0	1.0	12.0	4.0	3.0	3.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0
4	0.0	1.0	0.0	1.0	5.0	0.0	4.0	1.0	2.0	11.0	...	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0
...
665	0.0	2.0	0.0	4.0	2.0	2.0	5.0	5.0	0.0	0.0	...	0.0	0.0	1.0	0.0	0.0	1.0	1.0	2.0	3.0	0.0
666	0.0	1.0	0.0	2.0	0.0	0.0	9.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	3.0	0.0
667	0.0	2.0	0.0	4.0	1.0	1.0	2.0	5.0	3.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0
668	2.0	3.0	1.0	4.0	3.0	0.0	7.0	4.0	2.0	1.0	...	0.0	1.0	0.0	0.0	1.0	5.0	0.0	1.0	4.0	0.0
669	0.0	3.0	0.0	0.0	1.0	0.0	3.0	0.0	1.0	1.0	...	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0

670 rows x 50 columns

Collaborative Filtering 기법에 관한 연구

1. Matrix Factorization을 역이용해 평점 행렬을 생성



- ★ 형용사 Group을 이용해 만든 사용자(User), 장소(Place)별 Matrix를 구축한 뒤 행렬곱으로 평점 행렬을 만듦
- ★ 기존의 Collaborative Filtering은 평점 행렬 상에서 유사한 User를 골라 그 사람이 선호하는 Item을 추천해줌.

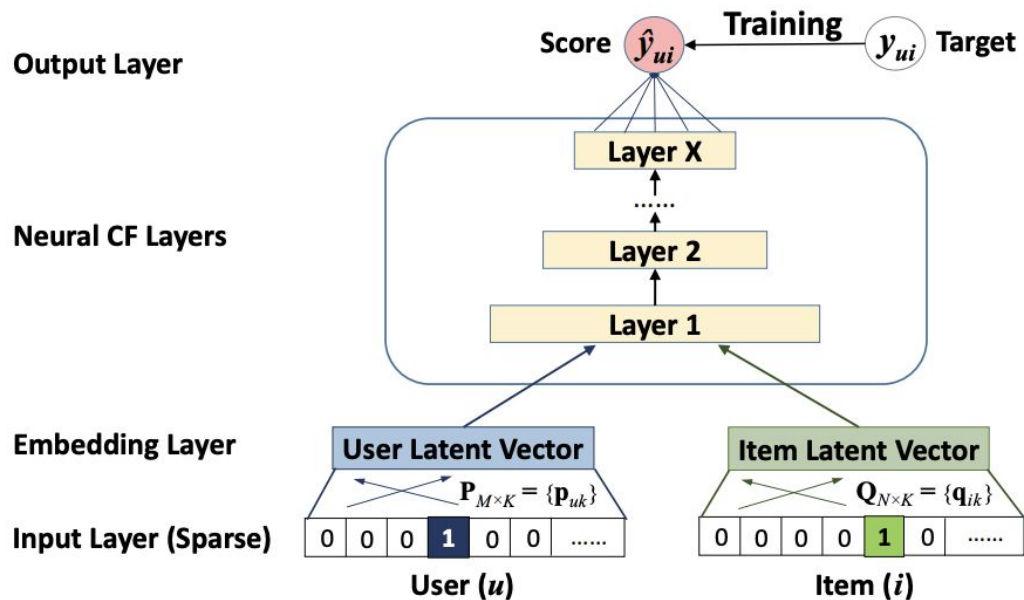
문제점

Sparse Matrix 문제 발생
다수의 데이터가 필요
유명한 항목 위주로 추천



Collaborative Filtering 기법에 관한 연구

2. 평점 행렬을 이용해 Neural Collaborative Filtering 실시



- 최근 추천 시스템의 트렌드는 유사도를 통해 최근접 이웃을 구하는 것 보다 잠재요인을 파악해 추천하는 것
- user latent factor와 item latent factor간의 상관관계를 표현하는데 MLP를 사용

Collaborative Filtering 기법에 관한 연구

3. 논문 참고 및 코드 검색

Neural Collaborative Filtering

Xiangnan He
National University of
Singapore, Singapore
xiangnanhe@gmail.com

Liqiang Nie
Shandong University
China
nieliqiang@gmail.com

Lizi Liao
National University of
Singapore, Singapore
liaolizi.llz@gmail.com

Xia Hu
Texas A&M University
USA
hu@cse.tamu.edu

Hanwang Zhang
Columbia University
USA
hanwangzhang@gmail.com

Tat-Seng Chua
National University of
Singapore, Singapore
dcscts@nus.edu.sg

ABSTRACT

In recent years, deep neural networks have yielded immense success on speech recognition, computer vision and natural language processing. However, the exploration of deep neural networks on recommender systems has received relatively less scrutiny. In this work, we strive to develop techniques based on neural networks to tackle the key problem in recommendation — collaborative filtering — on the basis of implicit feedback.

Although some recent work has employed deep learning for recommendation, they primarily used it to model auxiliary information, such as textual descriptions of items and acoustic features of musics. When it comes to model the key factor in collaborative filtering — the interaction between

1. INTRODUCTION

In the era of information explosion, recommender systems play a pivotal role in alleviating information overload, having been widely adopted by many online services, including E-commerce, online news and social media sites. The key to a personalized recommender system is in modelling users' preference on items based on their past interactions (*e.g.*, ratings and clicks), known as collaborative filtering [31, 46]. Among the various collaborative filtering techniques, matrix factorization (MF) [14, 21] is the most popular one, which projects users and items into a shared latent space, using a vector of latent features to represent a user or an item. Thereafter a user's interaction on an item is modelled as the inner product of their latent vectors.

hexiangnan / neural_collaborative_filtering

Code Issues 42 Pull requests 2 Actions Projects 0 Wiki

Neural Collaborative Filtering

deep-learning recommender-system collaborative-filtering

16 commits 1 branch 0 packages 0 releases

Branch: master New pull request

hexiangnan Merge pull request #34 from cakimuha/master	
Data	Change the name of Data
Pretrain	More detailed instructions.
Dataset.py	Add args to GMF and MLP
Dockerfile	Dockerfile
GMF.py	Add comments.
LICENSE	Add Apache License
MLP.py	Add comments.