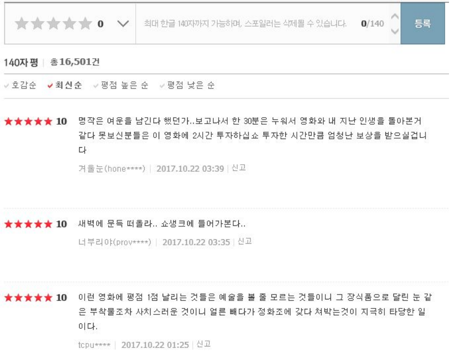
VOD RECOMMENDATION SYSTEM

설재완

Goal

How about others


★★★★★ 0
최대 한글 140자까지 가능하며, 스포일러는 삭제될 수 있습니다.
0/140
등록

140자 평 | 총 16,501건

✓ 호감순 ✓ 최신순 ✓ 평점 높은 순 ✓ 평점 낮은 순

★★★★★ 10 명작은 여운을 남긴다 했던가..보고나서 한 30분은 누워서 영화와 내 지난 인생을 돌아본거 같다 못보신분들은 이 영화에 2시간 투자하십쇼 투자한 시간만큼 엄청난 보상을 받으실겁니다

겨울눈(hone****) | 2017.10.22 03:39 | 신고

★★★★★ 10 새벽에 문득 떠올라.. 쇼생크에 들어가본다..

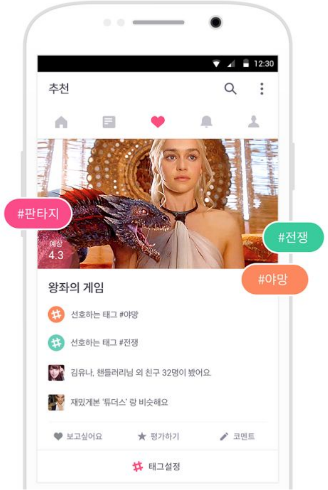
너부리야(prov****) | 2017.10.22 03:35 | 신고

★★★★★ 10 이런 영화에 평점 1점 날리는 것들은 예술을 볼 줄 모르는 것들이니 그 장식품으로 달린 눈 같은 부착물조차 사치스러운 것이니 얼른 빼다가 정화조에 갔다 쳐박는것이 지극히 타당한 일이다.

tcpu**** | 2017.10.22 01:25 | 신고

개인화 추천, 드라마까지

내게 맞는 영화뿐 아니라
이제 드라마도 쏙쏙



왕좌의 게임

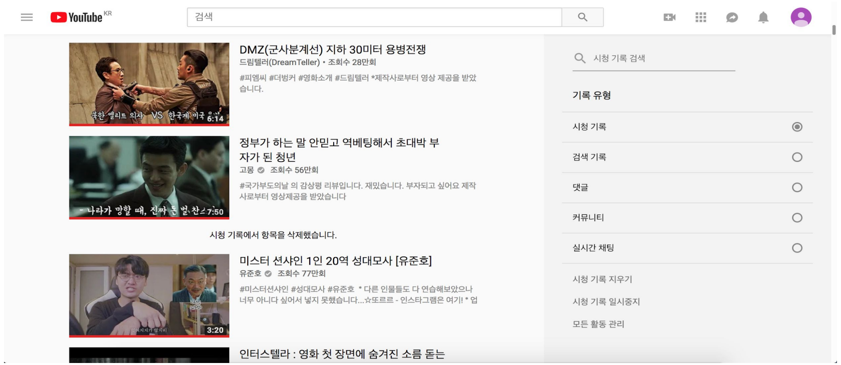
선호하는 태그 #야망

선호하는 태그 #전쟁

김유나, 챈들러리님 외 친구 32명이 봤어요.

재밌게본 '튜더스' 랑 비슷해요

No rating but history



DMZ(군사분계선) 지하 30미터 용병전쟁
드림텔러(DreamTeller) • 조회수 28만회
#피엠씨 #더벙커 #영화소개 #드림텔러 *제작사로부터 영상 제공을 받았습니다.

정부가 하는 말 안믿고 역베팅해서 초대박 부자가 된 청년
고몽 ✓ 조회수 56만회
#국가부도의날 의 감상평 리뷰입니다. 재밌습니다. 부자되고 싶어요 제작사로부터 영상제공을 받았습니다

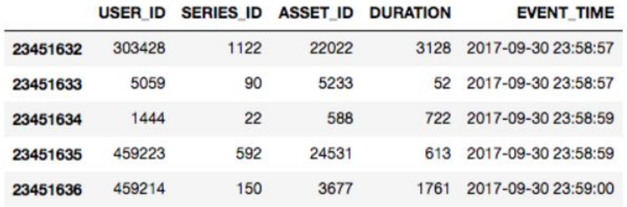
시청 기록에서 항목을 삭제했습니다.

미스터 션샤인 1인 20역 성대모사 [유준호]
유준호 ✓ 조회수 77만회
#미스터션샤인 #성대모사 #유준호 * 다른 인물들도 다 연습해보았으나 너무 아니다 싶어서 넣지 못했습니다...☆또르르 - 인스타그램은 여기! * 업

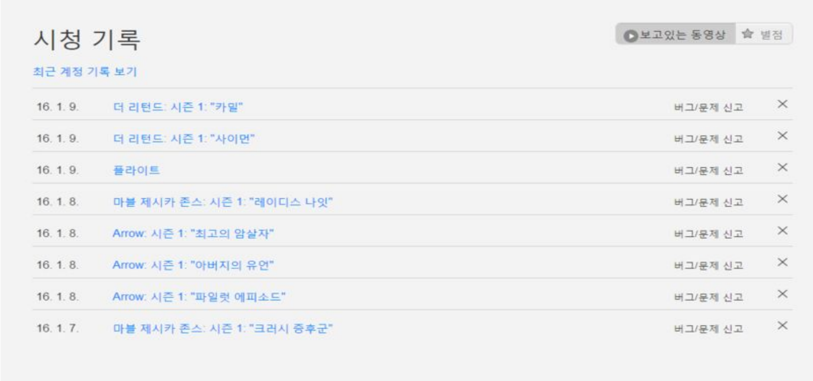
인터스텔라 : 영화 첫 장면에 숨겨진 소름 돋는

기록 유형

- 시청 기록 ●
- 검색 기록 ○
- 댓글 ○
- 커뮤니티 ○
- 실시간 채팅 ○



	USER_ID	SERIES_ID	ASSET_ID	DURATION	EVENT_TIME
23451632	303428	1122	22022	3128	2017-09-30 23:58:57
23451633	5059	90	5233	52	2017-09-30 23:58:57
23451634	1444	22	588	722	2017-09-30 23:58:59
23451635	459223	592	24531	613	2017-09-30 23:58:59
23451636	459214	150	3677	1761	2017-09-30 23:59:00

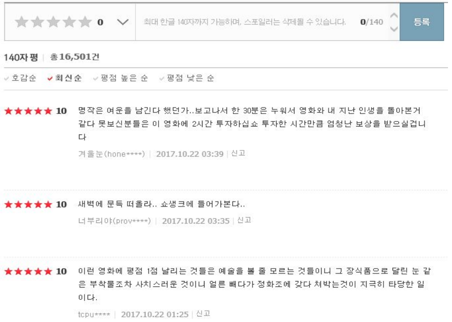


시청 기록

최근 계정 기록 보기

16. 1. 9.	더 리턴드: 시즌 1: "카밀"	버그/문제 신고	✕
16. 1. 9.	더 리턴드: 시즌 1: "사이먼"	버그/문제 신고	✕
16. 1. 9.	플라이트	버그/문제 신고	✕
16. 1. 8.	마블 제시카 존스: 시즌 1: "레이디스 나잇"	버그/문제 신고	✕
16. 1. 8.	Arrow: 시즌 1: "최고의 암살자"	버그/문제 신고	✕
16. 1. 8.	Arrow: 시즌 1: "아버지의 유언"	버그/문제 신고	✕
16. 1. 8.	Arrow: 시즌 1: "파일럿 에피소드"	버그/문제 신고	✕
16. 1. 7.	마블 제시카 존스: 시즌 1: "크러시 증후군"	버그/문제 신고	✕

Conversion From History To Rating



☆☆☆☆☆ 0 최대 한글 140자까지 가능하며, 스포일러는 삭제될 수 있습니다. 0/140 등록

140자 평 | 총 16,501건

✓ 호감순 ✓ 최신순 ✓ 평점 높은 순 ✓ 평점 낮은 순

★★★★★ 10 명작은 여운을 남긴다 했던가..보고나서 한 30분은 누워서 영화와 내 지난 인생을 돌아본거 같다 못보신분들은 이 영화에 2시간 투자하십쇼 투자한 시간만큼 엄청난 보상을 받으실겁니다

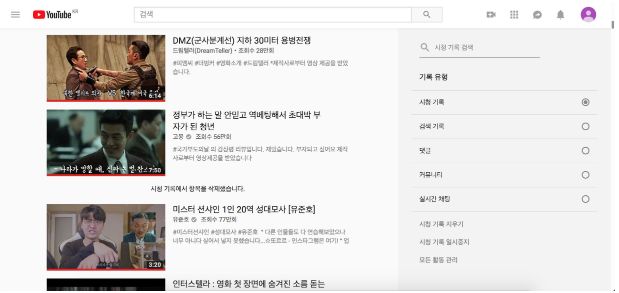
겨울눈(hone****) | 2017.10.22 03:39 | 신고

★★★★★ 10 새벽에 문득 떠올라.. 쇼생크에 들어가본다..

너부리야(prov****) | 2017.10.22 03:35 | 신고

★★★★★ 10 이런 영화에 평점 1점 날리는 것들은 예술을 볼 줄 모르는 것들이니 그 장식품으로 달린 눈 같은 부착물조차 사치스러운 것이니 얼른 빼다가 정화조에 갖다 쳐박는것이 지극히 타당한 일이다.

tcpu**** | 2017.10.22 01:25 | 신고



YouTube KR 검색

DMZ(군사분계선) 지하 30미터 용병전쟁

드림텔러(DreamTeller) • 조회수 28만회

#피엠씨 #더벙커 #영화소개 #드림텔러 *제작사로부터 영상 제공을 받았습니다.

정부가 하는 말 안믿고 역베팅해서 초대박 부자가 된 청년

고몽 조회수 56만회

#국가부도의날 의 감상평 리뷰입니다. 재밌습니다. 부자되고 싶어요 제작사로부터 영상제공을 받았습니다

시청 기록에서 항목을 삭제했습니다.

미스터 션샤인 1인 20역 성대모사 [유준호]

유준호 조회수 77만회

#미스터션샤인 #성대모사 #유준호 * 다른 인물들도 다 연습해보았으나 너무 아니다 싶어서 넣지 못했습니다...☆또르르 - 인스타그램은 여기! * 업

인터스텔라 : 영화 첫 장면에 숨겨진 소름 돋는

시청 기록 검색

기록 유형

시청 기록 ●

검색 기록 ○

댓글 ○

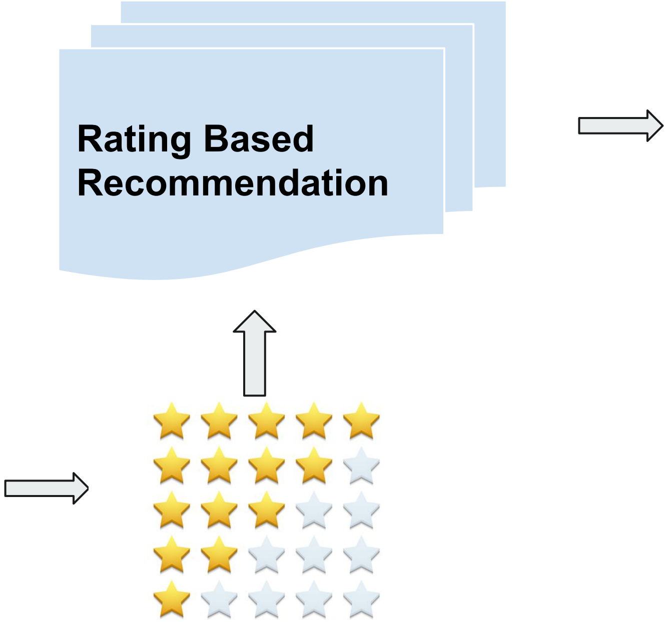
커뮤니티 ○

실시간 채팅 ○

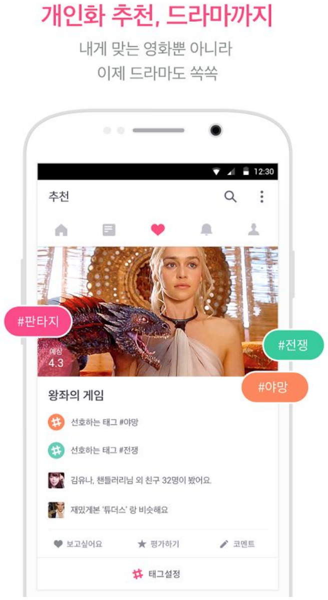
시청 기록 지우기

시청 기록 일시중지

모든 활동 관리



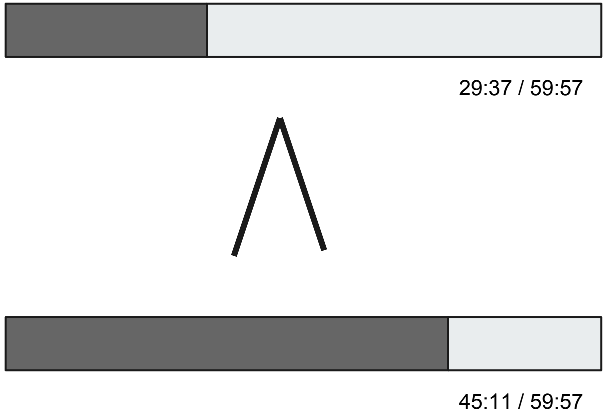
Rating Based Recommendation

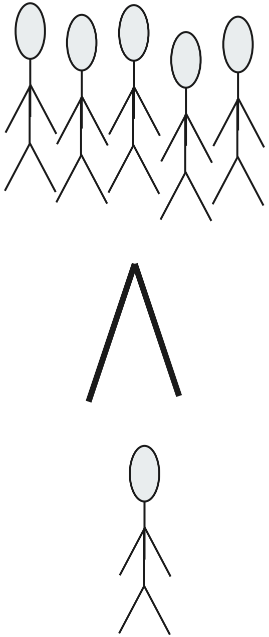


개인화 추천, 드라마까지

내게 맞는 영화뿐 아니라
이제 드라마도 쏙쏙

How to convert

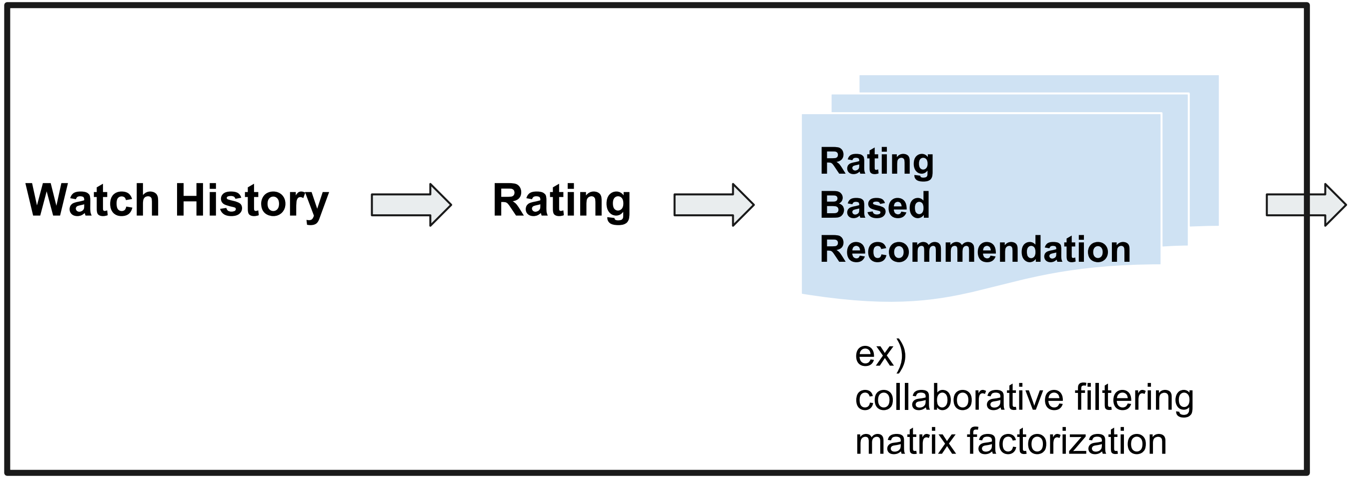




mean watching time = 30

watching time = 60

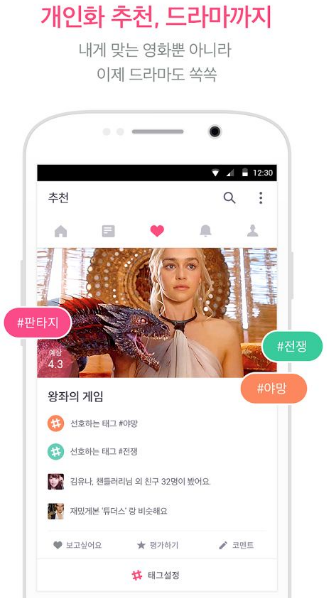
Recommendation Overview



Watch History → Rating → Rating Based Recommendation →

ex)
collaborative filtering
matrix factorization

history based recommendation



왕좌의 게임

선호하는 태그 #야망

선호하는 태그 #전쟁

김유나, 챈들러리님 외 친구 32명이 봤어요.

재밌게본 '튜더스' 랑 비슷해요

Implement - Rating Conversion

사람	영상	시청시간
갑	벡터맨	10
을	벡터맨	40
갑	벡터맨	10
갑	스칼라맨	30
병	스칼라맨	10

absolute time

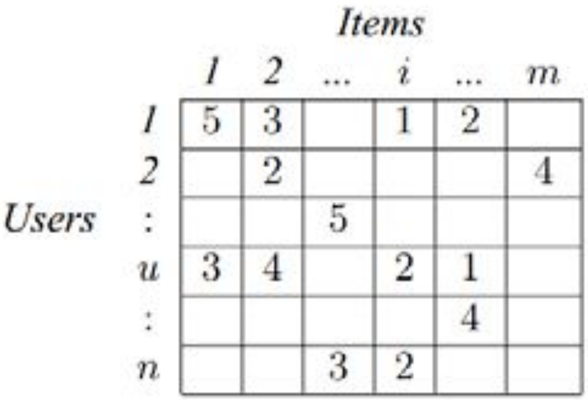
사람	영상	시청시간
갑	벡터맨	20
갑	스칼라맨	30
을	벡터맨	40
병	스칼라맨	10

relative time

	벡터맨	스칼라맨
갑	20 / 40 = 0.5	30 / 30 = 1
을	40 / 40 = 1	
병		10 / 30 = 0.33

user - asset matrix

Rating based recommendation ex)1



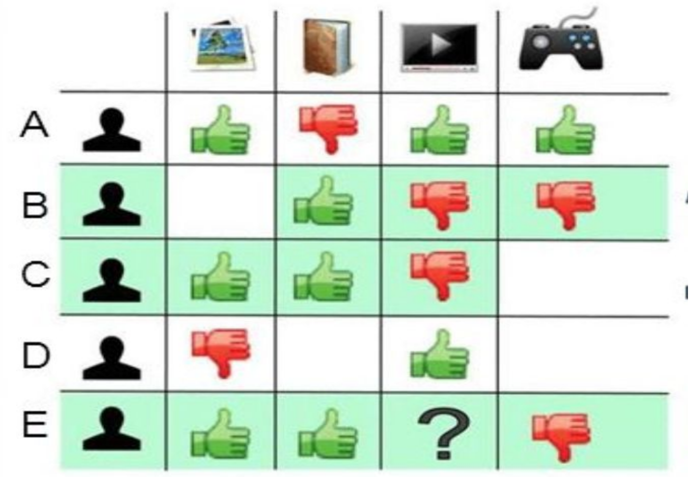
		Items					
		1	*2*	...	*i*	...	*m*
	1	5	3		1	2	
	2		2				4
Users	:			5			
	u	3	4		2	1	
	:					4	
	n			3	2		

baseline model

- consider each user and asset
- rui = mean + bu + bi
- also can be applied to another model

Rating based recommendation ex)2

collaborative filtering

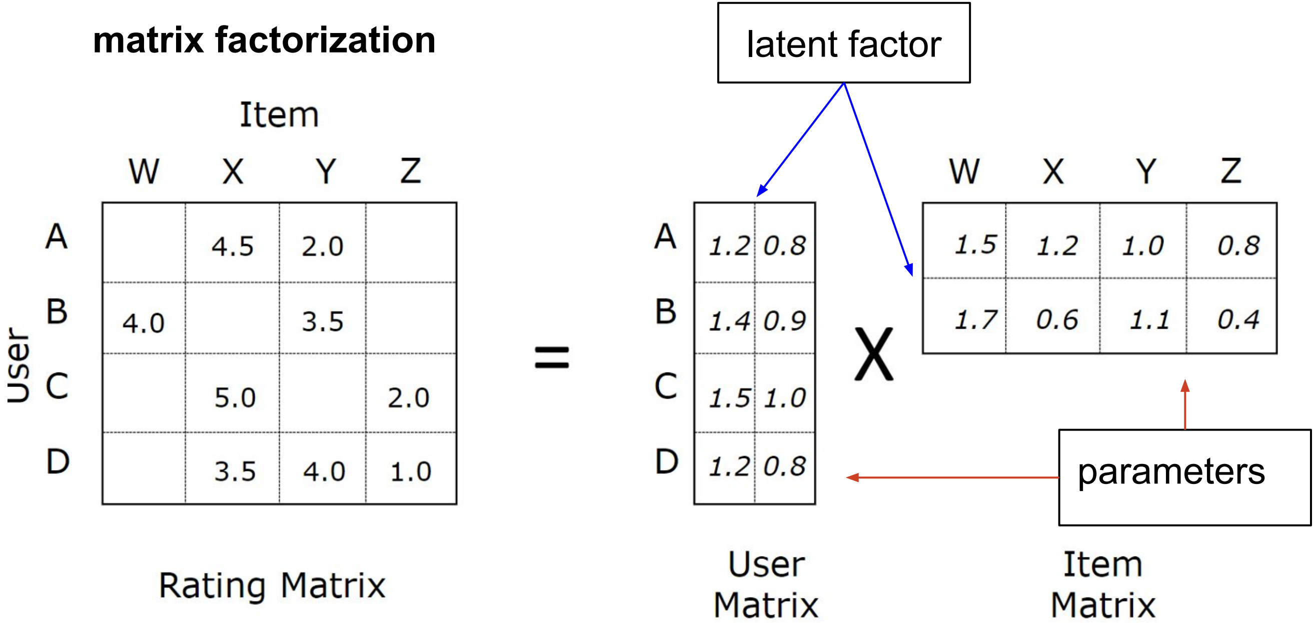


		🖼	📖	▶	🎮
A	👤	👍	👎	👍	👍
B	👤		👍	👎	👎
C	👤	👍	👍	👎	
D	👤	👎		👍	
E	👤	👍	👍	?	👎

- **Similarity**
  - **between item**
  - **between user**

Rating based recommendation ex)3

matrix factorization



Item

User	W	X	Y	Z
A		4.5	2.0	
B	4.0		3.5	
C		5.0		2.0
D		3.5	4.0	1.0

Rating Matrix

=

A	1.2	0.8
B	1.4	0.9
C	1.5	1.0
D	1.2	0.8

User Matrix

X

W	X	Y	Z
1.5	1.2	1.0	0.8
1.7	0.6	1.1	0.4

Item Matrix

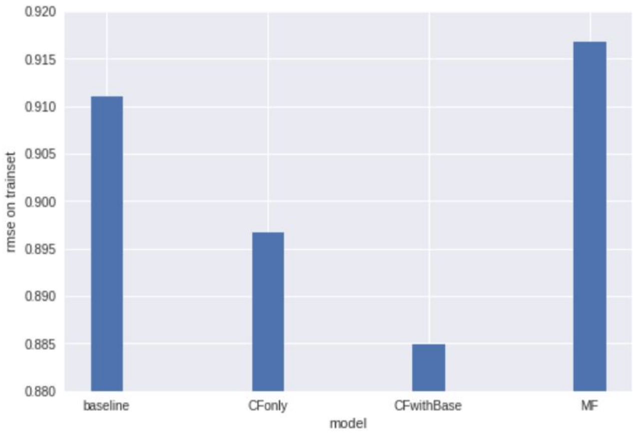
latent factor

parameters

Experiment

- 21,000,000 records
- 555,000 users
- 55,000 assets
- Compare and select model between
    - baseline
    - collaborative filtering only(CFonly)
    - collaborative filtering with baseline(CFwithBase)
    - matrix factorization(MF)
- Tain set: watch history of July, August
- Test set: watch history of September

Results



Final accuracy using CFwithBase: 47.24%

Limitation and future work

- If there were validation set, results will be better
- If there were good optimization technique(optimizer in ML/DL framework, fine tunning, etc), results will be better(It might not be trained well)
- If we can consider watching pattern(interval between each asset, consecutive assets, etc), results will be better
- If we use ‘genre’ and ‘series’ data well, results will be better
- If we pre-processed raw data well(there were so many histories of ‘뽀로로’), results will be better

Conclusion

- sutdy and implement basic ML concepts
    - use only numpy
    - gradient descnt, early stopping
- study and implement recommendation algorithm
    - collaborative filtering
    - matrix factorization
- read paper and implement
- problem solving life cycle
    - recognition -> background study -> design and implement -> evaluation -> result analysis

Thanks!

BACKUP SLIDES

Rating Conversion

```
-   Let arr[i, j] be rating of user i at asset j (after converting history to rating)
-   for jdx in range(len(asset)):
-       ratings = arr[:, j] # all ratings for asset j
-       median = sorted(ratings)[len(ratings) / 2] # median
-       arr[:, jdx] /= median # divide by median
```

Rating Conversion

- Relative time
  - t = (total watching time) / (running time)
  - can over-estimate for cases that short running time
- Absolute time
  - t = (total wathcing time)
  - can over-estimate for cases that long running time
- So, let's use (total watching time) / (mean or median watching time)
  - can consider other people(relative time) and pure wathcing time(absolute time) at once
  - median is better when extreme value exists