

Recommender Systems 1팀

하계 컨퍼런스 주제 발표

2022.07.05

김병준 김정우 김찬호 박정현

목차

- | 팀원 소개
- | 주제 및 선정 배경
- | Dataset
- | Models
- | 계획

팀원 소개

김병준 | 컴퓨터공학

김정우 | 경영학

김찬호 | 응용통계학

박정현 | 응용통계학

차애캐 & 애니메이션 추천



주제 및 선정 배경

“차애캐 추천과 그에 따른 애니메이션 추천”

영화 및 드라마의 경우 장르 불문 배우에 따라 작품을 소비하는 고객 (팬심 작용)



주제 및 선정 배경



금발
청소년
성급함
:



애니메이션의 경우 좋아하는 캐릭터가 있어도 다른 작품에서 찾는 것이 불가능

과거 해소할 수 없었던 needs를 충족시키는 새로운 방법 고민

'최애캐'의 외모와 성격 등 다양한 요소를 종합적으로 분석해 '차애캐'를 추천하는 방법론

이어서 차애캐가 등장하는 애니메이션 소개 후 해당 애니메이션과 유사한 애니메이션 추천

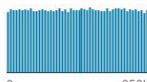

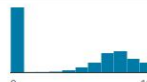


Dataset

kaggle 에 업로드 된 dataset 사용 예정

캐릭터 data는  anime-planet data 기반

애니메이션 data는  My AnimeList 기반

ID	Name	Alias	Gender	Hair Color
non identifiable randomly generated character id.	full name of this character.	another way to call the character.	gender of the this character.	hair color of this character.
 0 132k	131956 unique values	Unknown 89% Yuki 0% Other (14058) 11%	Male 51% Female 36% Other (17576) 13%	Unknown 34% Brown 18% Other (62110) 47%

user_id	anime_id	# rating	# watching_status	# watched_episodes
non identifiable randomly generated user id.	MyAnemlist ID of the anime.	score between 1 to 10 given by the user. 0 if the user didn't assign a score.	state ID from this anime in the anime list of this user.	numbers of episodes watched by the user.
 0 353k	 1 48.5k	 0 10	 0 55	 0 65.5k

Models

1. Contents-based Filtering
2. Collaborative Filtering (Item-based)
3. AutoRec
4. CDAE: Collaborative Denoising Auto-Encoder

Models

추천 시스템을 구현하기 위해 User와 Item으로 구성된 Data에 대한 파악이 필요

1. Contents-based Filtering 콘텐츠 기반 필터링

특정 item과 유사한 'contents'를 가진 다른 item 추천
최애캐와 여러모로 유사한 차애캐를 추천



2. Collaborative Filtering (Item-based) 아이템 기반 협업 필터링

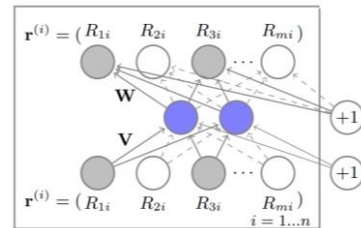
특정 item을 좋아한 users가 선호하는 또 다른 item 추천
'나루토'를 좋아한 사람들이 '강철의 연금술사'를 좋아하는 경향이 있다면,
'나루토'를 좋아하는 다른 사람에게 '강철의 연금술사' 추천



Models

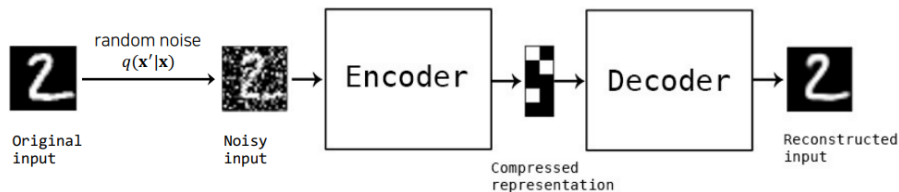
3. AutoRec

동일한 입력 출력 구조를 사용해 압축된
latent representation을 학습하는 비지도 학습 모델
애니메이션 평점 data에 masking 후 손실 함수를 최소화



4. CDAE: Collaborative Denoising Auto-Encoder

input에 noise를 준 noisy input을 사용하면 그 인풋의
latent representation을 학습해 원래의 input을 더 잘 복원하는 모델 구현
개별 user에 대해 item의 평점이 아닌 preference를 학습



계획

7월

week1

[주제 선정과 방법론 탐색]

week2

[데이터 전처리]

week3

[모델링 prototype 구축]

week4

[중간 발표]

8월

week1

[모델링 완성]

week2

[예선 제출 마감]

week3

[본선 발표 준비 및 발표]

*진행 상황에 따라 변동 가능

THOHI



감사합니다