

차애캐와 애니메이션 추천 시스템

Second Favorite Character & Animation
Recommender System

CUAI 5기 RecSys 10팀

김병준 김정우 김찬호 박정현



목차

서론
-연구 배경
-개요

콘텐츠 기반
추천 시스템
-차애캐 추천
-애니메이션 추천

협업 필터링
추천 시스템

결론
-한계점
-개선 방향



서론

연구 배경

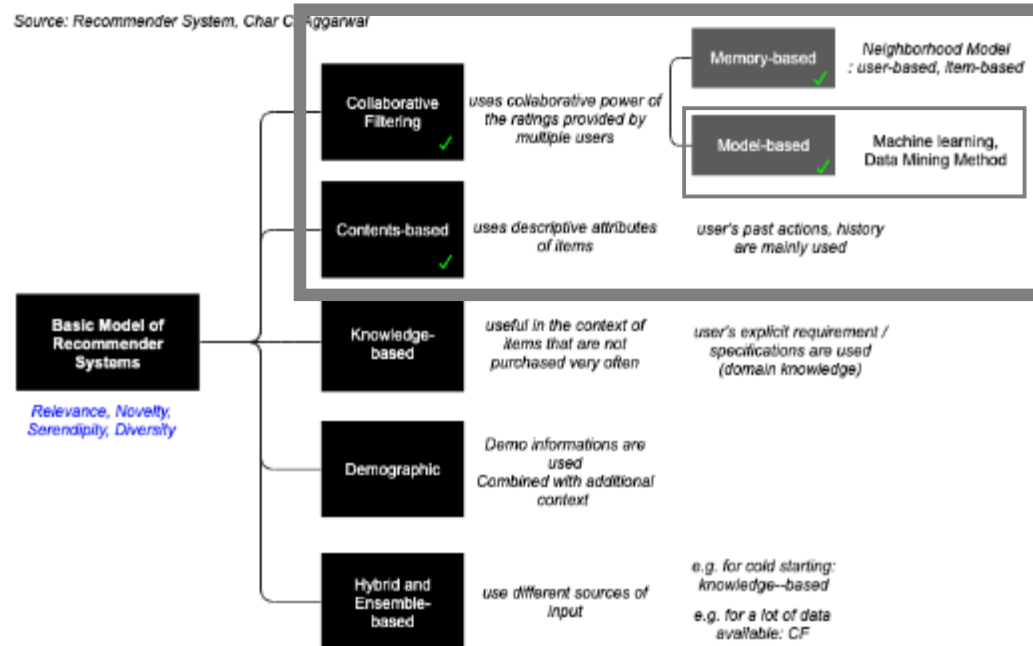
- OTT 플랫폼 산업과 통신 기술 발달
- 언제 어디서든 문화 콘텐츠를 소비할 수 있는 환경이 구축
- 여러 콘텐츠 중에서도 애니메이션에 대한 수요는 지속적으로 증가
- 애니메이션 플랫폼 ‘라프텔’은 ‘YouTube’에 이어 국내 기준 사용자 체류 시간 2위를 기록

동영상 플랫폼 순 방문자 수 TOP 10* (unit: 만 명)	동영상 플랫폼 평균 실행 횟수 TOP 10* (unit: 회)	동영상 플랫폼 평균 체류시간 TOP 10* (unit: 분)																																																																																																			
<table><tr><th colspan="3">APP</th></tr><tr><td>01</td><td>유튜브</td><td>2,887.1</td></tr><tr><td>02</td><td>네이버 밴드</td><td>1,585.6</td></tr><tr><td>03</td><td>인스타그램</td><td>1,105.8</td></tr><tr><td>04</td><td>페이스북</td><td>929.5</td></tr><tr><td>05</td><td>넷플릭스</td><td>342.5</td></tr><tr><td>06</td><td>웨이브</td><td>256.2</td></tr><tr><td>07</td><td>트위터</td><td>231.0</td></tr><tr><td>08</td><td>틱톡</td><td>209.4</td></tr><tr><td>09</td><td>U+ 모바일 yv</td><td>176.2</td></tr><tr><td>10</td><td>네이버TV</td><td>172.3</td></tr></table>	APP			01	유튜브	2,887.1	02	네이버 밴드	1,585.6	03	인스타그램	1,105.8	04	페이스북	929.5	05	넷플릭스	342.5	06	웨이브	256.2	07	트위터	231.0	08	틱톡	209.4	09	U+ 모바일 yv	176.2	10	네이버TV	172.3	<table><tr><th colspan="3">APP</th></tr><tr><td>01</td><td>트위터</td><td>290.7</td></tr><tr><td>02</td><td>페이스북</td><td>131.2</td></tr><tr><td>03</td><td>인스타그램</td><td>123.0</td></tr><tr><td>04</td><td>유튜브</td><td>111.7</td></tr><tr><td>05</td><td>틱톡</td><td>76.7</td></tr><tr><td>06</td><td>트위치</td><td>70.1</td></tr><tr><td>07</td><td>네이버밴드</td><td>62.6</td></tr><tr><td>08</td><td>아프리카TV</td><td>53.4</td></tr><tr><td>09</td><td>웨이브</td><td>32.5</td></tr><tr><td>10</td><td>넷플릭스</td><td>31.7</td></tr></table>	APP			01	트위터	290.7	02	페이스북	131.2	03	인스타그램	123.0	04	유튜브	111.7	05	틱톡	76.7	06	트위치	70.1	07	네이버밴드	62.6	08	아프리카TV	53.4	09	웨이브	32.5	10	넷플릭스	31.7	<table><tr><th colspan="3">APP</th></tr><tr><td>01</td><td>유튜브</td><td>1,464.5</td></tr><tr><td>02</td><td>트위터</td><td>966.7</td></tr><tr><td>03</td><td>트위치</td><td>713.3</td></tr><tr><td>04</td><td>웨이브</td><td>625.5</td></tr><tr><td>05</td><td>아프리카TV</td><td>571.2</td></tr><tr><td>06</td><td>페이스북</td><td>543.1</td></tr><tr><td>07</td><td>넷플릭스</td><td>494.8</td></tr><tr><td>08</td><td>틱톡</td><td>452.0</td></tr><tr><td>09</td><td>티빙</td><td>354.3</td></tr><tr><td>10</td><td>인스타그램</td><td>328.6</td></tr></table>	APP			01	유튜브	1,464.5	02	트위터	966.7	03	트위치	713.3	04	웨이브	625.5	05	아프리카TV	571.2	06	페이스북	543.1	07	넷플릭스	494.8	08	틱톡	452.0	09	티빙	354.3	10	인스타그램	328.6
APP																																																																																																					
01	유튜브	2,887.1																																																																																																			
02	네이버 밴드	1,585.6																																																																																																			
03	인스타그램	1,105.8																																																																																																			
04	페이스북	929.5																																																																																																			
05	넷플릭스	342.5																																																																																																			
06	웨이브	256.2																																																																																																			
07	트위터	231.0																																																																																																			
08	틱톡	209.4																																																																																																			
09	U+ 모바일 yv	176.2																																																																																																			
10	네이버TV	172.3																																																																																																			
APP																																																																																																					
01	트위터	290.7																																																																																																			
02	페이스북	131.2																																																																																																			
03	인스타그램	123.0																																																																																																			
04	유튜브	111.7																																																																																																			
05	틱톡	76.7																																																																																																			
06	트위치	70.1																																																																																																			
07	네이버밴드	62.6																																																																																																			
08	아프리카TV	53.4																																																																																																			
09	웨이브	32.5																																																																																																			
10	넷플릭스	31.7																																																																																																			
APP																																																																																																					
01	유튜브	1,464.5																																																																																																			
02	트위터	966.7																																																																																																			
03	트위치	713.3																																																																																																			
04	웨이브	625.5																																																																																																			
05	아프리카TV	571.2																																																																																																			
06	페이스북	543.1																																																																																																			
07	넷플릭스	494.8																																																																																																			
08	틱톡	452.0																																																																																																			
09	티빙	354.3																																																																																																			
10	인스타그램	328.6																																																																																																			
<table><tr><th colspan="3">WEB</th></tr><tr><td>01</td><td>유튜브</td><td>1,340.0</td></tr><tr><td>02</td><td>페이스북</td><td>564.1</td></tr><tr><td>03</td><td>네이버TV</td><td>436.6</td></tr><tr><td>04</td><td>카카오TV</td><td>282.5</td></tr><tr><td>05</td><td>인스타그램</td><td>267.4</td></tr><tr><td>06</td><td>트위터</td><td>187.5</td></tr><tr><td>07</td><td>넷플릭스</td><td>157.1</td></tr><tr><td>08</td><td>티빙</td><td>152.9</td></tr><tr><td>09</td><td>밴드</td><td>137.1</td></tr><tr><td>10</td><td>아프리카TV</td><td>126.4</td></tr></table>	WEB			01	유튜브	1,340.0	02	페이스북	564.1	03	네이버TV	436.6	04	카카오TV	282.5	05	인스타그램	267.4	06	트위터	187.5	07	넷플릭스	157.1	08	티빙	152.9	09	밴드	137.1	10	아프리카TV	126.4	<table><tr><th colspan="3">WEB</th></tr><tr><td>01</td><td>유튜브</td><td>99.6</td></tr><tr><td>02</td><td>트위터</td><td>87.3</td></tr><tr><td>03</td><td>라프텔</td><td>74.4</td></tr><tr><td>04</td><td>웨이브</td><td>50.9</td></tr><tr><td>05</td><td>아프리카TV</td><td>32.9</td></tr><tr><td>06</td><td>넷플릭스</td><td>32.8</td></tr><tr><td>07</td><td>트위치</td><td>23.7</td></tr><tr><td>08</td><td>밴드</td><td>21.0</td></tr><tr><td>09</td><td>티빙</td><td>18.7</td></tr><tr><td>10</td><td>인스타그램</td><td>18.0</td></tr></table>	WEB			01	유튜브	99.6	02	트위터	87.3	03	라프텔	74.4	04	웨이브	50.9	05	아프리카TV	32.9	06	넷플릭스	32.8	07	트위치	23.7	08	밴드	21.0	09	티빙	18.7	10	인스타그램	18.0	<table><tr><th colspan="3">WEB</th></tr><tr><td>01</td><td>유튜브</td><td>124.3</td></tr><tr><td>02</td><td>라프텔</td><td>78.8</td></tr><tr><td>03</td><td>트위터</td><td>60.9</td></tr><tr><td>04</td><td>웨이브</td><td>36.5</td></tr><tr><td>05</td><td>넷플릭스</td><td>26.9</td></tr><tr><td>06</td><td>트위치</td><td>24.4</td></tr><tr><td>07</td><td>아프리카TV</td><td>22.1</td></tr><tr><td>08</td><td>네이버TV</td><td>19.1</td></tr><tr><td>09</td><td>밴드</td><td>18.9</td></tr><tr><td>10</td><td>티빙</td><td>13.6</td></tr></table>	WEB			01	유튜브	124.3	02	라프텔	78.8	03	트위터	60.9	04	웨이브	36.5	05	넷플릭스	26.9	06	트위치	24.4	07	아프리카TV	22.1	08	네이버TV	19.1	09	밴드	18.9	10	티빙	13.6
WEB																																																																																																					
01	유튜브	1,340.0																																																																																																			
02	페이스북	564.1																																																																																																			
03	네이버TV	436.6																																																																																																			
04	카카오TV	282.5																																																																																																			
05	인스타그램	267.4																																																																																																			
06	트위터	187.5																																																																																																			
07	넷플릭스	157.1																																																																																																			
08	티빙	152.9																																																																																																			
09	밴드	137.1																																																																																																			
10	아프리카TV	126.4																																																																																																			
WEB																																																																																																					
01	유튜브	99.6																																																																																																			
02	트위터	87.3																																																																																																			
03	라프텔	74.4																																																																																																			
04	웨이브	50.9																																																																																																			
05	아프리카TV	32.9																																																																																																			
06	넷플릭스	32.8																																																																																																			
07	트위치	23.7																																																																																																			
08	밴드	21.0																																																																																																			
09	티빙	18.7																																																																																																			
10	인스타그램	18.0																																																																																																			
WEB																																																																																																					
01	유튜브	124.3																																																																																																			
02	라프텔	78.8																																																																																																			
03	트위터	60.9																																																																																																			
04	웨이브	36.5																																																																																																			
05	넷플릭스	26.9																																																																																																			
06	트위치	24.4																																																																																																			
07	아프리카TV	22.1																																																																																																			
08	네이버TV	19.1																																																																																																			
09	밴드	18.9																																																																																																			
10	티빙	13.6																																																																																																			



서론 개요

기술




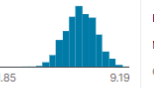
데이터셋



MyAnimeList 출처의
애니메이션 데이터



Anime-Planet 출처의
애니메이션 캐릭터 데이터

MAL_ID	Name	# Score	Genres
MyAnimeList ID of the anime.	full name of the anime	average score of the anime given from all users in MyAnimeList database.	comma separated list of genres for this anime.
	17558 unique values		Hentai 6%
			Music 4%
			Other (15803) 90%
1	Cowboy Bebop	8.78	Action, Adventure, Comedy, Drama, Sci-Fi, Space
5	Cowboy Bebop: Tengoku no Tobira	8.39	Action, Drama, Mystery, Sci-Fi, Space
6	Trigun	8.24	Action, Sci-Fi, Adventure, Comedy, Drama, Shounen



콘텐츠 기반 추천 시스템

Feature 기반 차애캐 추천

1) 'Anime-planet'에서 제공하는 캐릭터 dataset 전처리를 통해 필요한 정보만 추출

	ID	CharName	Gender	HairColor	Tags	Description	url	MAL_ID
0	0	L	Male	Black	Analytical,Barefoot,Detectives,EyeBags,SweetTooth	Secretive, meticulous and cunning, L's desire ...	https://www.anime-planet.com/characters/l-deat...	1535.0
1	98	Kakashi HATAKE	Male	Grey	Adults,Anti-GravityHair,Beastmasters,ExoticEye...	Kakashi is a jounin ninja from Konoha who is k...	https://www.anime-planet.com/characters/kakash...	34566.0
2	348	Light YAGAMI	Male	Brown	Analytical,BigEgo,Charismatic,GodComplex,Manip...	Armed with the power of Ryuk's Death Note, Lig...	https://www.anime-planet.com/characters/light-...	1535.0
3	279	Naruto UZUMAKI	Male	Blonde	Beastmasters,FaceMarkings,Headbands,Hot-Headed...	Naruto is a young ninja with a single dream: t...	https://www.anime-planet.com/characters/naruto...	35072.0
4	103	Edward ELRIC	Male	Blonde	Alchemists,ArmBlades,ArtificialLimbs,Braids,Gl...	Known as a prodigy and one of the strongest al...	https://www.anime-planet.com/characters/edward...	121.0



Feature 기반 차애캐 추천

```
1 characters.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 13888 entries, 0 to 13887  
Data columns (total 8 columns):  
#   Column          Non-Null Count  Dtype  
---  ---  
0   ID              13888 non-null  int64  
1   CharName        13888 non-null  object  
2   Gender          13888 non-null  object  
3   HairColor       13888 non-null  object  
4   Tags            13888 non-null  object  
5   Description     13888 non-null  object  
6   url             13888 non-null  object  
7   MAL_ID          13888 non-null  float64  
dtypes: float64(1), int64(1), object(6)  
memory usage: 868.1+ KB
```



콘텐츠 기반 추천 시스템

Feature 기반 차애캐 추천

* 추천에 사용되는 속성 3가지(Gender, HairColor, Tags)

Male, Female, Unknown,

▲ Gender에 등장하는 단어 뭉치 (3개)

Grey, Red, Orange, Purple, Black, Turquoise, Blue, Magenta, Pink, White, Blonde, Multicolored, Brown, Green,

▲ HairColor에 등장하는 단어 뭉치 (14개)

HistoricalFigures, Penguins, Directors, Circlets, FacialHair, Vocaloids, Bookworms, Androids, Catgirls, Nekomata, Landlords, Owls, Wolves, Archers, Airheads, Gamblers, Animators,

▲ Tags에 등장하는 단어 뭉치 (479개)



콘텐츠 기반 추천 시스템

Feature 기반 차애크 추천

2) Gender, HairColor, Tags를 bag_of_words라는 하나의 단어 집합으로 재생성.

ID	CharName	Gender	HairColor	Tags	Description	url	MAL_ID
0	0	L	Male	Black	Analytical,Barefoot,Detectives,EyeBags,SweetTooth	Secretive, meticulous and cunning, L's desire ...	https://www.anime-planet.com/characters/l-deat... 1535.0
1	98	Kakashi HATAKE	Male	Grey	Adults,Anti-GravityHair,Beastmasters,ExoticEye...	Kakashi is a jounin ninja from Konoha who is k...	https://www.anime-planet.com/characters/kakash... 34566.0
2	348	Light YAGAMI	Male	Brown	Analytical,BigEgo,Charismatic,GodComplex,Manip...	Armed with the power of Ryuk's Death Note, Lig...	https://www.anime-planet.com/characters/light-... 1535.0
3	279	Naruto UZUMAKI	Male	Blonde	Beastmasters,FaceMarkings,Headbands,Hot-Headed...	Naruto is a young ninja with a single dream: t...	https://www.anime-planet.com/characters/naruto... 35072.0
4	103	Edward ELRIC	Male	Blonde	Alchemists,ArmBlades,ArtificialLimbs,Braids,Gl...	Known as a prodigy and one of the strongest al...	https://www.anime-planet.com/characters/edward... 121.0



ID	CharName	Description	url	MAL_ID	bag_of_words	
0	0	L	Secretive, meticulous and cunning, L's desire ...	https://www.anime-planet.com/characters/l-deat...	1535.0	male black analytical barefoot detectives eyeb...
1	98	Kakashi HATAKE	Kakashi is a jounin ninja from Konoha who is k...	https://www.anime-planet.com/characters/kakash...	34566.0	male grey adults anti-gravityhair beastmasters...
2	348	Light YAGAMI	Armed with the power of Ryuk's Death Note, Lig...	https://www.anime-planet.com/characters/light-...	1535.0	male brown analytical bigego charismatic godco...
3	279	Naruto UZUMAKI	Naruto is a young ninja with a single dream: t...	https://www.anime-planet.com/characters/naruto...	35072.0	male blonde beastmasters facemarkings headband...
4	103	Edward ELRIC	Known as a prodigy and one of the strongest al...	https://www.anime-planet.com/characters/edward...	121.0	male blonde alchemists armblades artificiallim...



Feature 기반 차애캐 추천

3) CountVectorizer를 사용해서 각 단어에 고유한 정수 인덱스를 부여(단어 집합 생성) 한 후 각 인덱스의 위치에 단어 토큰의 등장 횟수를 기록한 벡터 생성

=> 13888x506 (13888은 캐릭터의 수, 506은 단어 집합의 수)

```
# count_matrix 생성
count = CountVectorizer()
count_matrix = count.fit_transform(characters['bag_of_words'])
```

```
array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 1, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       ...,
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 1, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0]])
```



Feature 기반 차애크 추천

4) Sklearn의 cosine_similarity를 이용하여 행렬 자기자신과의 코사인 유사도를 구해 유사도행렬 생성
=> 13888x13888 (13888은 캐릭터의 수)

```
# cosine_similarity matrix 생성
cosine_sim = cosine_similarity(count_matrix, count_matrix)
```

```
array([[1.          , 0.10910895, 0.25197632, ..., 0.21821789, 0.          ,
        0.          ],
       [0.10910895, 1.          , 0.09622504, ..., 0.33333333, 0.14433757,
        0.10206207],
       [0.25197632, 0.09622504, 1.          , ..., 0.38490018, 0.16666667,
        0.          ],
       ...,
       [0.21821789, 0.33333333, 0.38490018, ..., 1.          , 0.28867513,
        0.          ],
       [0.          , 0.14433757, 0.16666667, ..., 0.28867513, 1.          ,
        0.1767767 ],
       [0.          , 0.10206207, 0.          , ..., 0.          , 0.1767767 ,
        1.          ]])
```



Feature 기반 차애크 추천

* cosine_similarity 계산 식

$$\text{cosine similarity} = S_C(A, B) := \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$



콘텐츠 기반 추천 시스템

Feature 기반 차애캐 추천

4) 추천 받고자 하는 데이터의 인덱스 행을 내림차순으로 정렬하여 상위 k개의 추천 대상 index를 얻는다.



▲ Input Character

```
1 result = recommendations('Light YAGAMI')
```

Deishuu KAIKI
Gakuhou ASANO
Tsukasa SHISHIOU
Seiya KANIE
Johan LIEBERT
Shen Qing Qiu
Yu Wenzhou
Akushima
Fukusuke HIKYAKUYA
Ichiya SUZAKU
Sadasada TOKUGAWA
Kouichi SHIDO
Giovanni
Yamato HOUTSUIN
Suguru GETOU
Koreyuki TANI
Akoya's Father
Taichi MASHIMA
Koji YOSHIDA
Prime Minister Fujiwara



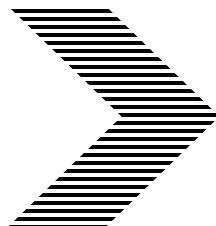
콘텐츠 기반 추천 시스템

Feature 기반 차애캐 추천

[Input]



▲ Light YAGAMI



[Output]



▲ Deishuu KAIKI



▲ Gakuhou ASANO



▲ Tsukasa SHISHIOU



▲ Seiya KANIE



Description 기반 차애캐 추천



1) 최애캐 설정

Description

Secretive, meticulous and cunning, L's
desire ...

Kakashi is a jounin ninja from Konoha
who is k...

Armed with the power of Ryuk's
Death Note, Lig...

2) 최애캐와
모든 애니메이션 캐릭터의
description을 비교

Sachiko YAGAMI
Heiji HATTORI
Kizaru
Cybersix
Kazunobu CHIBA
The Raven
Ritsu TAINAKA
Tsumugi KOTOBUKI
Dorry
Medusa GORGON

3) 가장 유사한 캐릭터를
차애캐로 추천



Description 기반 차애크 추천

TF-IDF는 전체 문서에서 자주 등장하는 단어의 중요도는 낮게 평가하는 반면, 특정 문서에서만 여러 차례 사용되는 단어의 중요도는 높게 평가합니다.

$$tf(t, d) = \frac{0.5 \times f(t, d)}{\max\{f(t, d) : w \in d\}}$$

$$idf(t, D) = \log \frac{|D|}{|\{d \in D : t \in d\}|}$$

$$tf \times idf$$



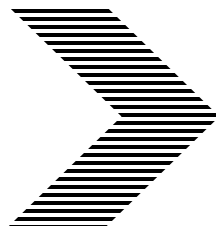
콘텐츠 기반 추천 시스템

Description 기반 차애캐 추천

[Input]



▲ Light YAGAMI



[Output]



▲ Sachiko YAGAMI



▲ Heiji HATTORI



▲ Kizaru



▲ Cybersix



Feature 기반 애니메이션 추천

```
anime_data.head()
```

	MAL_ID	Name	Genres	Type	Source	Year
0	1	Cowboy Bebop	Action,Adventure,Comedy,Drama,Sci-Fi,Space	TV	Original	1990.0
1	5	Cowboy Bebop: Tengoku no Tobira	Action,Drama,Mystery,Sci-Fi,Space	Movie	Original	2000.0
2	6	Trigun	Action,Sci-Fi,Adventure,Comedy,Drama,Shounen	TV	Manga	1990.0
3	7	Witch Hunter Robin	Action,Mystery,Police,Supernatural,Drama,Magic	TV	Original	2000.0
4	8	Bouken Ou Beet	Adventure,Fantasy,Shounen,Supernatural	TV	Manga	2000.0



Feature 기반 애니메이션 추천

```
anime_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 12181 entries, 0 to 12180  
Data columns (total 6 columns):  
#   Column      Non-Null Count  Dtype  
---  -  
0   MAL_ID      12181 non-null  int64  
1   Name        12181 non-null  object  
2   Genres      12181 non-null  object  
3   Type        12181 non-null  object  
4   Source      12181 non-null  object  
5   Year        11984 non-null  float64  
dtypes: float64(1), int64(1), object(4)  
memory usage: 571.1+ KB
```



Feature 기반 애니메이션 추천

1) Genre, Type, Source, Year를 CountVectorizer를 사용한 것과 같은 결과를 낼 수 있도록 속성에 등장하는 데이터를 읽어 단어뭉치 생성.

Police, Kids, Dementia, Mystery, Military, Unknown, Comedy, Thriller, School, Horror, Shounen Ai, Yaoi, Ecchi, Vampire, Romance, Harem, Historical, Shoujo, Sports, Seinen, SliceofLife, Mecha, Magic, Samurai, Game, Music, Josei, ShoujoAi, Shounen, Parody, Space, Fantasy, Demons, Cars, Adventure, MartialArts, Sci-Fi, Drama, SuperPower, Supernatural, Action, Psychological,

▲ genre에 등장하는 단어 뭉치

Music, ONA, Special, TV, Movie, OVA,

▲ type에 등장하는 단어 뭉치

Visual novel, Digital manga, Light novel, Game, Other, Music, Book, Picture book, Manga, Web manga, Original, Card game, Novel, 4-koma manga, Radio,

▲ source에 등장하는 단어 뭉치



콘텐츠 기반 추천 시스템

Feature 기반 애니메이션 추천

2) 단어뭉치에 존재하는 단어를 세며 one-hot vector 형식으로 변환하여 0과 1로 이루어진 행렬 생성.

```
score_full.shape  
(12181, 63)
```



Feature 기반 애니메이션 추천

- 3) Sklearn의 cosine_similarity를 이용하여 행렬 자기자신과의 코사인 유사도를 구해
(애니메이션 수 x 애니메이션 수) 크기의 유사도행렬 생성.

```
cosine_similar_full = cosine_similarity(score_full, score_full)
for i in range(len(cosine_similar_full)):
    cosine_similar_full[i,i] = 1.0
```

$$\text{cosine similarity} = S_C(A, B) := \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$



Feature 기반 애니메이션 추천

4) 추천 받고자 하는 데이터의 인덱스 행을 내림차순으로 정렬하여 상위 k개의 추천 대상 index를 얻는다.

```
def ani_rec_genre(df, cosine_similar_df, ani_name, top_n = 20):  
    ani_idx = anime_data[anime_data['Name'] == ani_name].index.values  
    sorted_df = cosine_similar_df[ani_idx]  
    sorted_df = sorted_df.sort_values(list(ani_idx), ascending = False)  
    top_n_idx = sorted_df.index[:top_n+1]  
    top_n_idx_without_input = list(set(top_n_idx) - set(ani_idx))  
    print(anime_data.loc[top_n_idx_without_input][['Name', 'MAL_ID']])  
    return list(anime_data.loc[top_n_idx_without_input]['MAL_ID'])
```



Feature 기반 애니메이션 추천



▲ Input Animation

```
result = ani_rec_genre(anime_data, cosine_similar_full_df, 'Naruto')
```

	Name	MAL_ID
4352	Medaka Box	11761
131	Shaman King	154
8010	Boruto: Naruto Next Generations	34566
3212	Dragon Ball Z: Atsumare! Gokuu World	6714
203	Rekka no Honoo	238
5518	Dragon Ball Kai (2014)	22777
8268	Boruto: Jump Festa 2016 Special	35072
1274	Katekyo Hitman Reborn!	1604
1366	Naruto: Shippuuden	1735
1178	D.Gray-man	1482
3035	Dragon Ball Kai	6033
233	Bleach	269
684	Dragon Ball Z	813
428	Ueki no Housoku	479
1329	Duel Masters	1685
11700	Shaman King (2021)	42205
824	Dragon Ball GT: Gokuu Gaiden! Yuuki no Akashi ...	987
6457	Dragon Ball Super	30694
762	Dragon Ball Z Movie 11: Super Senshi Gekiha!! ...	904
189	Dragon Ball	223



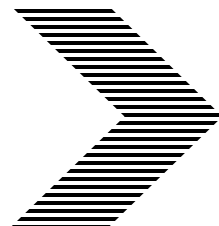
콘텐츠 기반 추천 시스템

Feature 기반 애니메이션 추천

[Input]



▲ Naruto



[Output]



▲ Boruto



▲ Naruto: Shippuuden



▲ Dragon Ball



▲ D.Gray-man



콘텐츠 기반 추천 시스템

Synopsis 기반 애니메이션 추천



1) 애니메이션
설정

2) 설정한 애니메이션과
모든 애니메이션의
Synopsis를 비교

synopsis

In the year 2071, humanity has colonized sever...

other day, another bounty—such is the life of ...

Vash the Stampede is the man with a \$\$60,000,0...

Boruto: Naruto Next Generations
Naruto: Shippuuden
Boruto: Naruto the Movie
Naruto: Shippuuden Movie 6 - Road to Ninja
Naruto SD: Rock Lee no Seishun Full-Power Ninden
Naruto: Takigakure no Shitou - Ore ga Eiyuu Da...
The Last: Naruto the Movie
Naruto: Shippuuden Movie 4 - The Lost Tower
Naruto: Shippuuden Movie 2 - Kizuna
Naruto: Akaki Yotsuba no Clover wo Sagase
Naruto: Shippuuden Movie 1

3) 가장 유사한
애니메이션 추천



Synopsis 기반 애니메이션 추천

TF-IDF는 전체 문서에서 자주 등장하는 단어의 중요도는 낮게 평가하는 반면, 특정 문서에서만 여러 차례 사용되는 단어의 중요도는 높게 평가합니다.

$$tf(t, d) = \frac{0.5 \times f(t, d)}{\max\{f(t, d) : w \in d\}}$$

$$idf(t, D) = \log \frac{|D|}{|\{d \in D : t \in d\}|}$$

$$tf \times idf$$



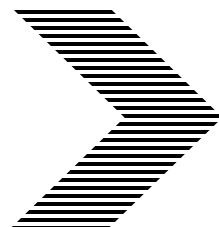
콘텐츠 기반 추천 시스템

Synopsis 기반 애니메이션 추천

[Input]



▲ Naruto



[Output]



▲ Boruto: Naruto Next Generation



▲ Naruto: Shippuuden



▲ Boruto: Naruto the Movie



▲ The Last: Naruto the Movie



협업 필터링 추천 시스템

Collaborative Filtering

“나와 성향이 비슷한 친구들은 어떤 애니메이션을 찾아볼까?”

=> 많은 사용자들로부터 얻은 기호정보(taste information)에 따라 사용자들의
관심사들을 자동적으로 예측하게 해주는 방법

The logo for the 'surprise' Python library, featuring the word 'surprise' in a white, lowercase, sans-serif font. The letter 'i' is replaced by a white exclamation mark.

A Python scikit for
recommender systems.



협업 필터링 추천 시스템

데이터셋

	user_id	anime_id	rating
0	478	3613	7
1	478	5462	6
2	478	33023	7
3	478	176	5
4	478	5332	6
...
4354363	352930	671	7
4354364	352930	1017	8
4354365	352930	10465	6
4354366	352930	10845	6
4354367	352930	14045	7

Surprise를 활용하여 추천시스템을 구현하고자 할 때, '사용자 - 아이템 - 평점'으로 데이터 가공 필요



협업 필터링 추천 시스템 알고리즘

▼ Surprise 라이브러리에서 제공하는 여러 알고리즘

random_pred.NormalPredictor	Algorithm predicting a random rating based on the distribution of the training set, which is assumed to be normal.
baseline_only.BaselineOnly	Algorithm predicting the baseline estimate for given user and item.
knns.KNNBasic	A basic collaborative filtering algorithm.
knns.KNNWithMeans	A basic collaborative filtering algorithm, taking into account the mean ratings of each user.
knns.KNNWithZScore	A basic collaborative filtering algorithm, taking into account
knns.KNNBaseline	A basic collaborative filtering algorithm taking into account a <i>baseline</i> rating.
matrix_factorization.SVD	The famous <i>SVD</i> algorithm, as popularized by Simon Funk during the Netflix Prize. When baselines are not used, this is equivalent to Probabilistic Matrix Factorization.
matrix_factorization.SVDpp	The <i>SVD++</i> algorithm, an extension of SVD taking into account implicit ratings.
matrix_factorization.NMF	A collaborative filtering algorithm based on Non-negative Matrix Factorization.
slope_one.SlopeOne	A simple yet accurate collaborative filtering algorithm.
co_clustering.CoClustering	A collaborative filtering algorithm based on co-clustering.

ALGORITHM	TEST_RMSE	TEST_MAE
SVD	1.0564	0.7801
SLOPEONE	1.2064	0.9022
NMF	1.8523	1.5859
KNNBASELINE	1.1582	0.8625
KNNBASIC	1.1920	0.8840
KNNWITHMEANS	1.1685	0.8729
KNNWITHZSCORE	1.1764	0.8708
BASELINEONLY	1.2085	0.9038
COCLUSTERING	1.1924	0.8917

▲ 협업 필터링 알고리즘 성능 교차 검증 결과



협업 필터링 추천 시스템

SVD 알고리즘

$$\begin{array}{c} \boxed{A} \\ t \times d \end{array} = \begin{array}{c} \boxed{U} \\ t \times m \end{array} \begin{array}{c} \boxed{\Sigma} \\ m \times m \end{array} \begin{array}{c} \boxed{V^T} \\ m \times d \end{array} \approx \begin{array}{c} \boxed{U_k} \\ t \times k \end{array} \begin{array}{c} \boxed{\Sigma_k} \\ k \times k \end{array} \begin{array}{c} \boxed{V_k^T} \\ k \times d \end{array} = \begin{array}{c} \boxed{A_k} \\ t \times d \end{array}$$

A : 사용자-아이템 평점 행렬
U : 아이템 행렬
 Σ : 특잇값 행렬
V : 사용자 행렬
 A_k : 예측 행렬

SVD(Singular Value Decomposition, 특이값 분해)는 사용자-아이템 평점 행렬을 3개의 행렬(사용자 행렬(V) X 특잇값 행렬(Σ) X 아이템 행렬(U))로 분해해서 이를 학습시키고 이 3개의 행렬로 원래의 행렬을 재현하여 해당 사용자가 평가하지 않은 아이템에 대한 평점을 예측하는 기법이다.



SVD 파라미터 조정

- **n_epochs**: SGD의 반복 횟수 지정
- **n_factors**: SVD의 잠재 요인 K의 크기 지정
- **lr_all**: 학습률
- **reg_all**: 규제 텀 계수

최적화할 파라미터들을 딕셔너리 형태로 지정.

```
param_grid = {'n_epochs': [17, 18, 19], 'n_factors': [3000, 4500, 6000] }
```

CV를 3개 폴드 세트로 지정, 성능 평가는 rmse, mse 로 수행 하도록 GridSearchCV 구성

```
gs = GridSearchCV(SVD, param_grid, measures=['rmse', 'mae'], cv=5, n_jobs=-1)
gs.fit(data)
```

최고 RMSE Evaluation 점수와 그때의 하이퍼 파라미터

```
print(gs.best_score['rmse'])
print(gs.best_params['rmse'])
```



협업 필터링 추천 시스템

추천 시스템 구현

```
def recomm_anime_by_surprise(algo, user_id, unseen_animes, top_n=10):
    # 알고리즘 객체의 predict() 메서드를 평점이 없는 애니메이션에 반복 수행한 후 결과를 list 객체로 저장
    predictions = [algo.predict(str(user_id), str(anime_id)) for anime_id in unseen_animes]

    # predictions list 객체는 surprise의 Predictions 객체를 원소로 가지고 있음.
    # [Prediction(uid='9', iid='1', est=3.69), Prediction(uid='9', iid='2', est=2.98),...]
    # 이를 est 값으로 정렬하기 위해서 아래의 sortkey_est 함수를 정의함.
    # sortkey_est 함수는 list 객체의 sort() 함수의 키 값으로 사용되어 정렬 수행.
    def sortkey_est(pred):
        return pred.est

    # sortkey_est( ) 반환값의 내림 차순으로 정렬 수행하고 top_n개의 최상위 값 추출.
    predictions.sort(key=sortkey_est, reverse=True)
    top_predictions= predictions[:top_n]

    # top_n으로 추출된 애니메이션의 정보 추출. 애니메이션 아이디, 추천 예상 평점, 제목 추출
    top_anime_ids = [int(pred.iid) for pred in top_predictions]
    top_anime_rating = [pred.est for pred in top_predictions]
    top_anime_names = anime[anime.MAL_ID.isin(top_anime_ids)][['title']]
    top_anime_preds = [(id, title, rating) for id, title, rating in zip(top_anime_ids, top_anime_names, top_anime_rating)]

    return top_anime_preds

unseen_animes = get_unseen_animelist(9)
top_anime_preds = recomm_anime_by_surprise(algo, 9, unseen_animes, top_n=10)
print('##### Top-10 추천 애니메이션 리스트 #####')

for top_anime in top_anime_preds:
    print(top_anime[1], ":", top_anime[2])
```

- 1) 사용할 알고리즘과 추천 대상자인 User의 id를 입력 받음.
- 2) User_id에 해당하는 사용자가 이미 평점을 매긴, 즉 이미 시청한 애니메이션을 제외.
- 3) 입력된 알고리즘에 의해 사용자가 평점을 매기지 않은 애니메이션의 평점을 예측.
- 4) 예측된 평점 중 가장 높은 평점을 받은 상위 N개의 애니메이션을 추천.



협업 필터링 추천 시스템

추천 결과

Katekyo Hitman Reborn! : 9.083834449003131

Mononoke : 8.968436672746423

Major S5 : 8.896757860526067

Hunter x Hunter (2011) : 8.856850018649558

Mahou Shoujo Madoka★Magica Movie 3: Hangyaku no Monogatari : 8.789397204110024

Gintama* : 8.768553224869489

Kimi no Na wa. : 8.759830850041032

Gintama. : 8.715619723113676

Re:Creators : 8.70757873411669

Tsurezure Children : 8.691663300444697

▲ 사용자 아이디(user_id)가 478인 사용자에게 추천된 애니메이션 10개



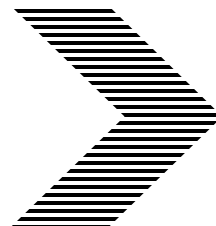
협업 필터링 추천 시스템

추천 결과

[Input]

User ID 478

평점 매긴 애니메이션 수 1204
추천 대상 애니메이션 수 10977
전체 애니메이션 수 12181



[Output]



▲ Kateyo Hitman
Reborn!



▲ Mononoke



▲ Major S5



▲ Hunter X Hunter

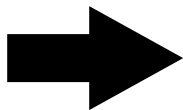


결론

한계점과 개선 방향- Feature 기반 추천 시스템

“전혀 연관성이 없을 것으로 생각되는 후보가 여럿 등장”

1. 모든 속성이 동일한 가중치를 가진다.
2. 데이터셋이 가지는 여러 feature에 등장하는 단어의 수가 일정하지 않다.



속성을 Vectorize할 때 속성에 따라 가중치를 다르게 설정하거나
데이터셋에 양질의 feature를 추가.

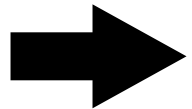


결론

한계점과 개선 방향- Description, Sysnopsis 기반 추천 시스템

“추천의 다양성 부족”

고유 명사에 부여하는 가중치가 크기 때문에 차애캐로는 같은 애니메이션에 등장한 캐릭터 위주로, 애니메이션으로는 비슷한 시리즈의 애니메이션을 위주로 추천한다.



같은 애니메이션에 등장한 캐릭터나 같은 시리즈의 애니메이션을 추천 대상에서 배제.





감사합니다.

Thank You

