추천 시스템

Recommender System

협업 필터링
Collaborative filtering

컨텐츠 기반 필터링
Content based filtering

하이브리드 필터링
Hybrid filtering

User based Collaborative filtering

어떤 User와 가장 비슷한가?

→ Smilarity (유사도) 이용

User based Collaborative filtering

USER 3 에게 추천할 영화는?

Cosine Similarity 사용

좋아요 : 1
그 외 : 0

	Antman	Avengers	Spiderman	Titanic	Gatsby
USER 1	1	1	1	0	0
USER 2	0	0	0	1	1
USER 3	1	1	0	0	0
USER 4	0	0	0	0	1

User based Collaborative filtering

	Antman	Avengers	Spiderman	Titanic	Gatsby
USER 1	1	1	1	0	0
USER 2	0	0	0	1	1
USER 3	1	1	0	0	0
USER 4	0	0	0	0	1

$$\frac{(1*1)+(1*1)+(0*1)+(0*0)+(0*0)}{\sqrt{(1^2+1^2+0^2+0^2+0^2}*\sqrt{(1^2+1^2+1^2+0^2+0^2}} = \frac{2}{\sqrt{2}*\sqrt{3}} = 0.82$$

	Antman	Avengers	Spiderman	Titanic	Gatsby
USER 1	1	1	1	0	0
USER 2	0	0	0	1	1
USER 3	1	1	0	0	0
USER 4	0	0	0	0	1

$$\frac{(1*0)+(1*0)+(0*0)+(0*1)+(0*1)}{\sqrt{(1^2+1^2+0^2+0^2+0^2}*\sqrt{(0^2+0^2+0^2+1^2+1^2}} = \frac{0}{\sqrt{2}*\sqrt{2}} = 0$$

	Antman	Avengers	Spiderman	Titanic	Gatsby
USER 1	1	1	1	0	0
USER 2	0	0	0	1	1
USER 3	1	1	0	0	0
USER 4	0	0	0	0	1

$$\frac{(1*1)+(1*1)+(0*0)+(0*0)+(0*0)}{\sqrt{(1^2+1^2+0^2+0^2+0^2}*\sqrt{(1^2+1^2+0^2+0^2+0^2}} = \frac{1}{\sqrt{2}*\sqrt{2}} = 1$$

	Antman	Avengers	Spiderman	Titanic	Gatsby
USER 1	1	1	1	0	0
USER 2	0	0	0	1	1
USER 3	1	1	0	0	0
USER 4	0	0	0	0	1

$$\frac{(1*0)+(1*0)+(0*0)+(0*0)+(0*1)}{\sqrt{(1^2+1^2+0^2+0^2+0^2}*\sqrt{(0^2+0^2+0^2+0^2+1^2}} = \frac{0}{\sqrt{2}*\sqrt{1}} = 0$$

Cosine similarity : 1 (3&3) > 0.82 (3&1) > 0 (3&2) (3&4)

User based Collaborative filtering

USER 3 에게 추천할 영화는?

Cosine Similarity 사용

좋아요 : 1
그 외 : 0

	Antman	Avengers	Spiderman	Titanic	Gatsby
USER 1	1	1	1	0	0
USER 2	0	0	0	1	1
USER 3	1	1	0	0	0
USER 4	0	0	0	0	1

USER 1 (0.82)

➔ USER 3가 보지 않은 영화 중 USER 1이 본 영화 추천 : SPIDERMAN

Item based Collaborative filtering

USER 4 에게 추천할 영화는?

Cosine Similarity 사용

좋아요 : 1
그 외 : 0

	USER 1	USER 2	USER 3	USER 4
Antman	1	0	1	0
Avengers	1	0	1	0
Spiderman	1	0	0	0
Titanic	0	1	0	0
Gatsby	0	1	0	1

Item based Collaborative filtering

	USER 1	USER 2	USER 3	USER 4
Antman	1	0	1	0
Avengers	1	0	1	0
Spiderman	1	0	0	0
Titanic	0	1	0	0
Gatsby	0	1	0	1

$$\frac{(0*1)+(1*0)+(0*1)+(1*0)}{\sqrt{(0^2+1^2+0^2+1^2}*\sqrt{(1^2+0^2+1^2+0^2}}=\frac{0}{\sqrt{2}*\sqrt{2}}=0$$

	USER 1	USER 2	USER 3	USER 4
Antman	1	0	1	0
Avengers	1	0	1	0
Spiderman	1	0	0	0
Titanic	0	1	0	0
Gatsby	0	1	0	1

$$\frac{(0*1)+(1*0)+(0*1)+(1*0)}{\sqrt{(0^2+1^2+0^2+1^2}*\sqrt{(1^2+0^2+1^2+0^2}}=\frac{0}{\sqrt{2}*\sqrt{2}}=0$$

	USER 1	USER 2	USER 3	USER 4
Antman	1	0	1	0
Avengers	1	0	1	0
Spiderman	1	0	0	0
Titanic	0	1	0	0
Gatsby	0	1	0	1

$$\frac{(0*1)+(1*0)+(0*0)+(1*0)}{\sqrt{(0^2+1^2+0^2+1^2}*\sqrt{(1^2+0^2+0^2+0^2}}=\frac{0}{\sqrt{2}*\sqrt{1}}=0$$

	USER 1	USER 2	USER 3	USER 4
Antman	1	0	1	0
Avengers	1	0	1	0
Spiderman	1	0	0	0
Titanic	0	1	0	0
Gatsby	0	1	0	1

$$\frac{(0*0)+(1*1)+(0*0)+(1*0)}{\sqrt{(0^2+1^2+0^2+1^2}*\sqrt{(0^2+1^2+0^2+0^2}}=\frac{1}{\sqrt{2}*\sqrt{1}}=0.71$$

	USER 1	USER 2	USER 3	USER 4
Antman	1	0	1	0
Avengers	1	0	1	0
Spiderman	1	0	0	0
Titanic	0	1	0	0
Gatsby	0	1	0	1

$$\frac{(0*0)+(1*1)+(0*0)+(1*1)}{\sqrt{(0^2+1^2+0^2+1^2}*\sqrt{(0^2+1^2+0^2+1^2}}=\frac{2}{\sqrt{2}*\sqrt{2}}=1$$

Cosine similarity : 1 (G&G) > 0.71 (G&T) > 0 (G&A) (G&A) (G&S)

Item based Collaborative filtering

USER 4 에게 추천할 영화는? Cosine Similarity 사용

좋아요 : 1
그 외 : 0

	USER 1	USER 2	USER 3	USER 4
Antman	1	0	1	0
Avengers	1	0	1	0
Spiderman	1	0	0	0
Titanic	0	1	0	0
Gatsby	0	1	0	1

Titanic (0.71)

➔ USER 4가 보지 않은 영화 중 GATSBY와 유사한 영화 추천 : TITANIC

COLD START 문제

새로운 영화 (NEW AVENGERS) 는 누구에게 추천?

Cosine Similarity 사용

좋아요 : 1
그 외 : 0

	Avengers	Spiderman	Titanic	Gatsby
USER 1	1	1	0	0
USER 2	0	0	1	1
USER 3	1	0	0	0
USER 4	0	0	0	1

NEW AVENGERS

COLD START 문제

데이터가 없는 시작 상태에서 제대로 동작하지 않는 시스템

➔ CONTENT BASED FILTERING 사용

Content based filtering

새로운 영화 (NEW AVENGERS) 는 누구에게 추천?

Cosine Similarity 사용

O : 1
X : 0

	Iron_Man	Captain_America	Spiderman	Leonardo_Dicaprio	Based_on_the_true_story
New Avengers	1	1	1	0	0
Avengers	1	1	0	0	0
Spiderman	1	0	1	0	0
Titanic	0	0	0	1	1
Gatsby	0	0	0	1	0

Content based filtering

	Iron_Man	Captain_America	Spiderman	Leonardo_Dicaprio	Based_on_the_true_story
New Avengers	1	1	1	0	0
Avengers	1	1	0	0	0
Spiderman	1	0	1	0	0
Titanic	0	0	0	1	1
Gatsby	0	0	0	1	0

$$\frac{(1*1)+(1*1)+(1*0)+(0*0)+(0*0)}{\sqrt{(1^2+1^2+1^2+0^2+0^2}*\sqrt{(1^2+1^2+0^2+0^2+0^2}} = \frac{2}{\sqrt{3}*\sqrt{2}} = 0.82$$

	Iron_Man	Captain_America	Spiderman	Leonardo_Dicaprio	Based_on_the_true_story
New Avengers	1	1	1	0	0
Avengers	1	1	0	0	0
Spiderman	1	0	1	0	0
Titanic	0	0	0	1	1
Gatsby	0	0	0	1	0

$$\frac{(1*1)+(1*0)+(1*1)+(0*0)+(0*0)}{\sqrt{(1^2+1^2+1^2+0^2+0^2}*\sqrt{(1^2+0^2+1^2+0^2+0^2}} = \frac{2}{\sqrt{3}*\sqrt{2}} = 0.82$$

	Iron_Man	Captain_America	Spiderman	Leonardo_Dicaprio	Based_on_the_true_story
New Avengers	1	1	1	0	0
Avengers	1	1	0	0	0
Spiderman	1	0	1	0	0
Titanic	0	0	0	1	1
Gatsby	0	0	0	1	0

$$\frac{(1*0)+(1*0)+(1*0)+(0*1)+(0*1)}{\sqrt{(1^2+1^2+1^2+0^2+0^2}*\sqrt{(0^2+0^2+0^2+1^2+1^2}} = \frac{0}{\sqrt{3}*\sqrt{2}} = 0$$

	Iron_Man	Captain_America	Spiderman	Leonardo_Dicaprio	Based_on_the_true_story
New Avengers	1	1	1	0	0
Avengers	1	1	0	0	0
Spiderman	1	0	1	0	0
Titanic	0	0	0	1	1
Gatsby	0	0	0	1	0

$$\frac{(1*0)+(1*0)+(1*0)+(0*1)+(0*0)}{\sqrt{(1^2+1^2+1^2+0^2+0^2}*\sqrt{(0^2+0^2+0^2+1^2+0^2}} = \frac{0}{\sqrt{3}*\sqrt{1}} = 0$$

Cosine similarity : 0.82 (N&A) (N&S) > 0 (N&T) (N&G)

Content based filtering

새로운 영화 (NEW AVENGERS) 는 누구에게 추천?

Cosine Similarity 사용

O : 1
X : 0

	Iron_Man	Captain_America	Spiderman	Leonardo_Dicaprio	Based_on_the_true_story
New Avengers	1	1	1	0	0
Avengers	1	1	0	0	0
Spiderman	1	0	1	0	0
Titanic	0	0	0	1	1
Gatsby	0	0	0	1	0

Avengers, Spiderman (0.82)

➔ New Avengers 영화는 Avengers 와 Spiderman 을 좋아하는 사람에게 추천

참조

유튜브 Minsuk Heo : 추천 시스템 기본 - (콜라보레이티브 필터링, 컨텐트 베이스 필터링)
https://www.youtube.com/watch?v=_YndKkun2Sw&list=PLk-YqoQ-G1VSRxweRK8_OSQD17lR4_5iG&index=50