

5주차 과제

Collaborative filtering



담당교수님: 윤장혁 교수님

과목명: Data Analytics

이름: 박민성

학번: 201611145

전공: 산업공학과

제출일: 2020.04.19

$r_{u,i}$ 을 분석한 후, User X에게 item을 추천해준다.

data_week5.txt 를 정리한 결과이다. User(1-10)의 item에 대한 선호도 평가이다.

	Item 1	2	3	4	5	6	7	8	9	10
user 1		4		3		2		1	2	3
2	5	2	1	3	2		1			
3	1	1		2	3	5		2	4	4
4			4	5		3		3		
5	2	1			4	3		1	1	2
6	2		4	2		1	3	5	1	2
7	1	3		2	5		2		1	4
8		1	1	2	3	3		3	1	
9	3		2		1		4		5	
10	1	2	1	3	4			2	1	1

User X의 평가이다.

	item 1	2	3	4	5	6	7	8	9	10
user x		2	1		3			4	3	

우선, User X와 User 1-10 중에서 어떤 User가 X와 비슷한지 판별하기 위해서

Cosine-based similarity between users x and y

$$sim(x, y) \approx \cos(\vec{x}, \vec{y}) = \frac{\sum_{i \in I_{xy}} r_{x,i} r_{y,i}}{\sqrt{\sum_{i \in I_{xy}} r_{x,i}^2} \sqrt{\sum_{i \in I_{xy}} r_{y,i}^2}}$$

Cosine-based similarity을 활용한다.

1	2	3	4	5	6	7	8	9	10
0.7294	0.98	0.9181	0.7761	0.7815	0.8171	0.8649	0.9435	0.8377	0.8793

<표 1>

계산 Tool은 엑셀을 활용하였다.

<표 1> 계산 예시

User 1과 user x의 similarity

$$(4 \times 2 + 1 \times 4 + 2 \times 3) / \{ (4^2 + 1^2 + 2^2) * (2^2 + 4^2 + 3^2) \}^{0.5}$$

$$= 0.7294$$

Rating prediction of an item을 진행한다.

각 User들의 점수를 주는 tendency를 고려하기 위해 각 선호도에서 각 개인의 선호도의 평균을 빼준다.

	item 1	2	3	4	5	6	7	8	9	10	Similarity	평균
1		1.50		0.50		-0.50		-1.50	-0.50	0.50	0.7294	2.50
2	2.67	-0.33	-1.33	0.67	-0.33		-1.33				0.98	2.33
3	-1.75	-1.75		-0.75	0.25	2.25		-0.75	1.25	1.25	0.9181	2.75
4			0.25	1.25		-0.75		-0.75			0.7761	3.75
5	0.00	-1.00			2.00	1.00		-1.00	-1.00	0.00	0.7815	2.00
6	-0.50		1.50	-0.50		-1.50	0.50	2.50	-1.50	-0.50	0.8171	2.50
7	-1.57	0.43		-0.57	2.43		-0.57		-1.57	1.43	0.8649	2.57
8		-1	-1	0	1	1		1	-1		0.9435	2.00
9	0.00		-1.00		-2.00		1.00		2.00		0.8377	3.00
10	-0.88	0.12	-0.88	1.12	2.12			0.12	-0.88	-0.88	0.8793	1.88

<표 2>

<표 2> 계산 예시

User 1 item 2의 4(평가)-2.5(평균)=1.5

	item 1	2	3	4	5	6	7	8	9	10
1	0.0000	1.0941	0.0000	0.3647	0.0000	-0.3647	0.0000	-1.0941	-0.3647	0.3647
2	1.9475	-0.2407	-0.9701	0.4887	-0.2407	0.0000	-0.9701	0.0000	0.0000	0.0000
3	-1.2765	-1.2765	0.0000	-0.5471	0.1824	1.6412	0.0000	-0.5471	0.9118	0.9118
4	0.0000	0.0000	0.1824	0.9118	0.0000	-0.5471	0.0000	-0.5471	0.0000	0.0000
5	0.0000	-0.7294	0.0000	0.0000	1.4588	0.7294	0.0000	-0.7294	-0.7294	0.0000
6	-0.3647	0.0000	1.0941	-0.3647	0.0000	-1.0941	0.3647	1.8235	-1.0941	-0.3647
7	-1.1452	0.3136	0.0000	-0.4158	1.7724	0.0000	-0.4158	0.0000	-1.1452	1.0430
8	0.0000	-0.7294	-0.7294	0.0000	0.7294	0.7294	0.0000	0.7294	-0.7294	0.0000
9	0.0000	0.0000	-0.7294	0.0000	-1.4588	0.0000	0.7294	0.0000	1.4588	0.0000
10	-0.6419	0.0875	-0.6419	0.8169	1.5463	0.0000	0.0000	0.0875	-0.6419	-0.6419
sum	-1.4807	-1.4807	-1.7943	1.2546	3.9898	1.0941	-0.2918	-0.2772	-2.3341	1.3129
$k \times \text{sum}$	-0.1737	-0.1737	-0.2105	0.1472	0.4680	0.1283	-0.0342	-0.0325	-0.2738	0.1540
$r_{u,i}$	2.4263	2.4263	2.3895	2.7472	3.0680	2.7283	2.5658	2.5675	2.3262	2.7540

<표 3>

$\bar{r}_u = 2.6$ (User x의 선호도 평균)

$$r_{u,i} = \bar{r}_u + k \sum_{u' \in U} \text{sim}(u, u') \times (r_{u',i} - \bar{r}_{u'})$$

where k is a normalizing factor defined as $k = \frac{1}{\sum_{u' \in U} \text{sim}(u, u')}$

여기서 k 값은 similarity의 합의 역수인

$$1/(0.7294+0.98+0.9181+0.7761+0.7815+0.8171+0.8649+0.9435+0.8377+0.8793) = 0.1173 \text{ 이다.}$$

<표 3>의 계산 예시

Item 1의 계산 예시

$$2.6 + (0 \cdot 0.7294 + 2.67 \cdot 0.98 + -1.75 \cdot 0.9181 + 0 \cdot 0.7761 + 0 \cdot 0.7815 + -0.5 \cdot 0.8171 + -1.57 \cdot 0.8649 + 0 \cdot 0.9435 + 0 \cdot 0.8377 + -0.88 \cdot 0.8793) \cdot 0.1173 = 2.4263$$

User X가 사용하지 않았던 제품(1,4,6,7,10) 중에 추천해준다.

$r_{u,i}$ 값에 따라 item 10, item 4, item 6, item 7, item 1순으로 User X에게 item을 추천해준다.