

Recommender Systems

Chap. 2

Neighborhood-Based Collaborative Filtering

Neighborhood-based algorithms

- User-based collaborative filtering
 - 타겟 A와 유사한 유저를 기반으로 A에게 추천한다. (peer group)
- Item-based collaborative filtering
 - Item B와 가장 유사한 Item 집합 S를 결정한 후 유저 A를 기반으로 S에서 B를 예측한다.

Key Properties of Ratings Matrices

- Ratings를 정의하는 다양한 방법들
 - Continuous ratings: like or dislike
 - Interval-based ratings: Numerical values로 표현
 - Ordinal ratings: 아주 싫다, 싫다, 보통, 좋다, 아주 좋다
 - Binary ratings: positive or negative
 - Unary ratings: 페이스북의 좋아요 버튼
- Long-tail property
 - Item의 빈도가 높을 수록 낮은 Rating을 가지는 분포

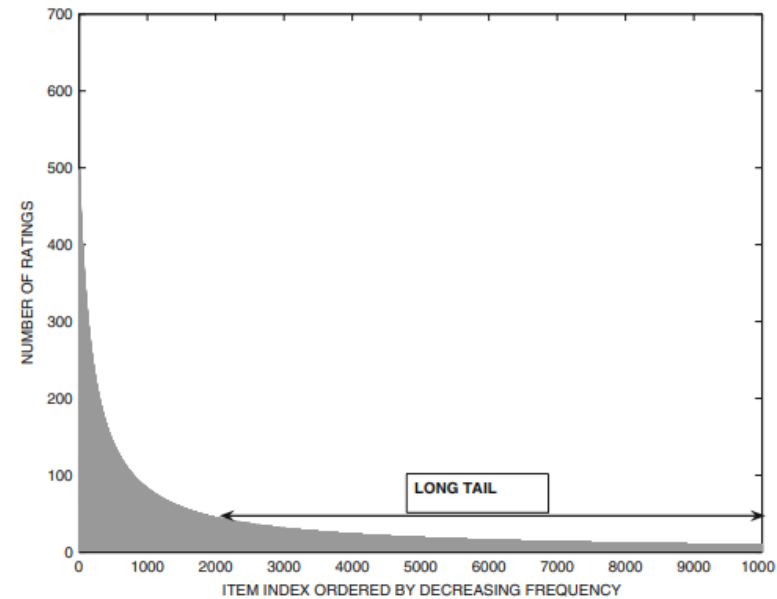


Figure 2.1: The long tail of rating frequencies

Predicting Ratings

User-based Models

Table 2.1: User-user similarity computation between user 3 and other users

Item-Id \Rightarrow	1	2	3	4	5	6	Mean Rating	Cosine($i, 3$) (user-user)	Pearson($i, 3$) (user-user)
User-Id \Downarrow									
1	7	6	7	4	5	4	5.5	0.956	0.894
2	6	7	?	4	3	4	4.8	0.981	0.939
3	?	3	3	1	1	?	2	1.0	1.0
4	1	2	2	3	3	4	2.5	0.789	-1.0
5	1	?	1	2	3	3	2	0.645	-0.817

Item-based Models

Table 2.2: Ratings matrix of Table 2.1 with mean-centering for adjusted cosine similarity computation among items. The adjusted cosine similarities of items 1 and 6 with other items are shown in the last two rows.

Item-Id \Rightarrow	1	2	3	4	5	6
User-Id \Downarrow						
1	1.5	0.5	1.5	-1.5	-0.5	-1.5
2	1.2	2.2	?	-0.8	-1.8	-0.8
3	?	1	1	-1	-1	?
4	-1.5	-0.5	-0.5	0.5	0.5	1.5
5	-1	?	-1	0	1	1
Cosine(1, j) (item-item)	1	0.735	0.912	-0.848	-0.813	-0.990
Cosine(6, j) (item-item)	-0.990	-0.622	-0.912	0.829	0.730	1

User-Based

- Rating이 계산되는 target user와 유사한 user를 식별하기 위한 방법
- Target user i를 찾기위해 다른 모든 user들이 계산된다.
- 따라서, Similarity functio이 정의되어야한다.

$$\mu_u = \frac{\sum_{k \in I_u} r_{uk}}{|I_u|} \quad \forall u \in \{1 \dots m\}$$

$$\text{Sim}(u, v) = \text{Pearson}(u, v) = \frac{\sum_{k \in I_u \cap I_v} (r_{uk} - \mu_u) \cdot (r_{vk} - \mu_v)}{\sqrt{\sum_{k \in I_u \cap I_v} (r_{uk} - \mu_u)^2} \cdot \sqrt{\sum_{k \in I_u \cap I_v} (r_{vk} - \mu_v)^2}}$$

User-Based

- 서로 다른 user들이 서로 다른 scales의 ratings를 줄 수 있기 때문에 mean-centered rating이 필요하다.

$$s_{uj} = r_{uj} - \mu_u \quad \forall u \in \{1 \dots m\}$$

- Prediction function

$$\hat{r}_{uj} = \mu_u + \frac{\sum_{v \in P_u(j)} \text{Sim}(u, v) \cdot s_{vj}}{\sum_{v \in P_u(j)} |\text{Sim}(u, v)|} = \mu_u + \frac{\sum_{v \in P_u(j)} \text{Sim}(u, v) \cdot (r_{vj} - \mu_v)}{\sum_{v \in P_u(j)} |\text{Sim}(u, v)|}$$

- Standardized rating

$$z_{uj} = \frac{r_{uj} - \mu_u}{\sigma_u} = \frac{s_{uj}}{\sigma_u}$$

$$\hat{r}_{uj} = \mu_u + \sigma_u \frac{\sum_{v \in P_u(j)} \text{Sim}(u, v) \cdot z_{vj}}{\sum_{v \in P_u(j)} |\text{Sim}(u, v)|}$$

Item-Based

- Item-based model에서 peer group들은 users가 아닌 items로 구성되어야한다.
- 아이템들 간의 유사도가 계산되어야한다.

Item-Based

Adjusted cosine similarity

$$\text{AdjustedCosine}(i, j) = \frac{\sum_{u \in U_i \cap U_j} s_{ui} \cdot s_{uj}}{\sqrt{\sum_{u \in U_i \cap U_j} s_{ui}^2} \cdot \sqrt{\sum_{u \in U_i \cap U_j} s_{uj}^2}}$$

Predicted rating

$$\hat{r}_{ut} = \frac{\sum_{j \in Q_t(u)} \text{AdjustedCosine}(j, t) \cdot r_{uj}}{\sum_{j \in Q_t(u)} |\text{AdjustedCosine}(j, t)|}$$

Efficient Implement and Computational Complexity

- 기본 접근: 가능한 모든 유저-아이템 쌍에 대한 평점을 예측하고 순위를 매김
- Offline phase: user-user (or item-item) 유사도 값과 peer group⁰이 계산된다.
- Online phase: Offline phase에서 계산된 값을 사용해 예측
- Offline phase의 계산 시간이 더 많이 걸린다.

Comparing User and Item Based

User-Based

- 더 다양한 추천을 제공
- 구체적인 근거 제시 불가

Item-Based

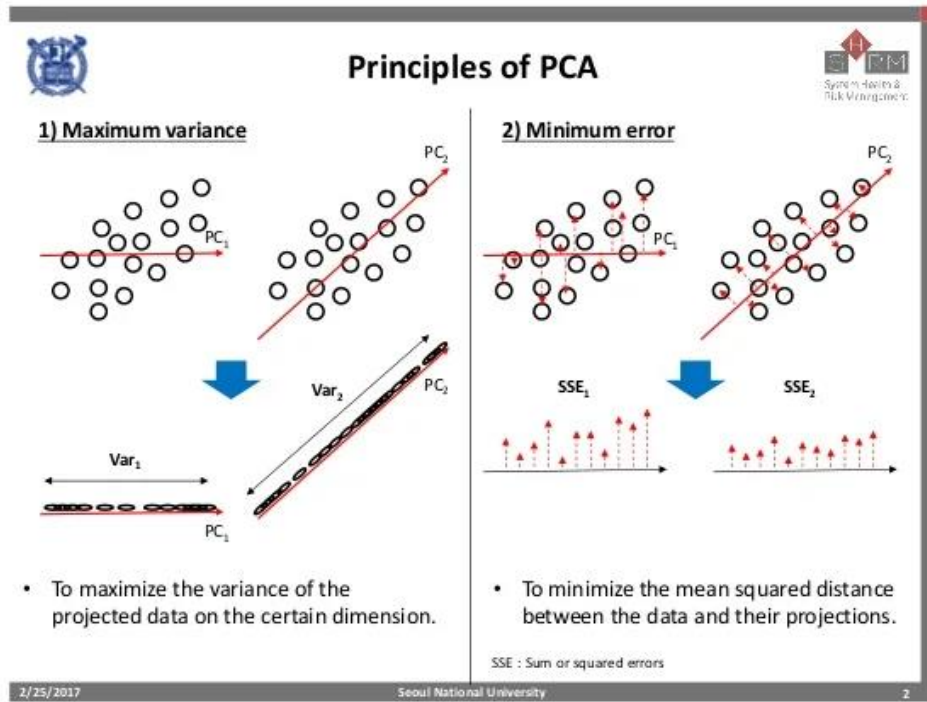
- Rating 변화에 더 안정적임
- 구체적인 근거 제시

Clustering

- Offline 단계의 복잡성이 가장 큰 문제
- 클러스터링을 통해 가장 근접한 이웃 단계를 클러스터링 단계로 대체함
- 클러스터가 세분화될 수록 정확도는 조금 떨어짐

Dimension Reduction

PCA



SVD

$$\begin{matrix}
 \begin{matrix} \text{Matrix} \\ m \times n \end{matrix} & = &
 \begin{matrix} \text{Matrix} \\ m \times m \end{matrix} &
 \begin{matrix} \text{Matrix} \\ m \times n \end{matrix} &
 \begin{matrix} \text{Matrix} \\ n \times n \end{matrix} \\
 \mathbf{M} & = & \mathbf{U} & \mathbf{\Sigma} & \mathbf{V}^* \\
 m \times n & & m \times m & m \times n & n \times n
 \end{matrix}$$

$$\begin{matrix}
 \begin{matrix} \text{Matrix} \\ m \times m \end{matrix} & & \begin{matrix} \text{Matrix} \\ m \times m \end{matrix} & = & \begin{matrix} \text{Matrix} \\ m \times m \end{matrix} \\
 \mathbf{U} & & \mathbf{U}^* & = & \mathbf{I}_m
 \end{matrix}$$

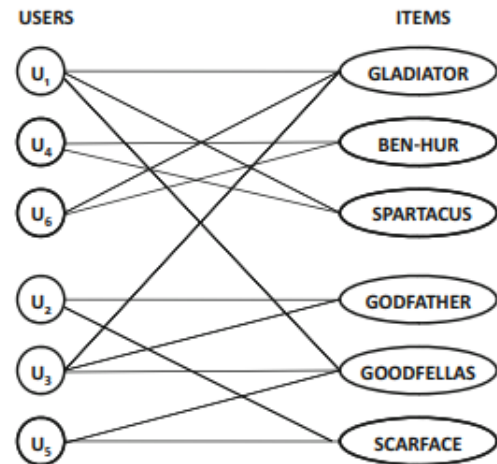
$$\begin{matrix}
 \begin{matrix} \text{Matrix} \\ m \times n \end{matrix} & & \begin{matrix} \text{Matrix} \\ m \times n \end{matrix} & = & \begin{matrix} \text{Matrix} \\ m \times n \end{matrix} \\
 \mathbf{V} & & \mathbf{V}^* & = & \mathbf{I}_n
 \end{matrix}$$

Graph Models

User-Item

	GLADIATOR	GODFATHER	BEN-HUR	GOODFELLAS	SCARFACE	SPARTACUS
U_1	1			5		2
U_2		5			4	
U_3	5	3		1		
U_4			3			4
U_5				3	5	
U_6	5		4			

(a) Ratings matrix



(b) User-item graphs of specified ratings

Figure 2.3: A ratings matrix and corresponding user-item graph

User-User

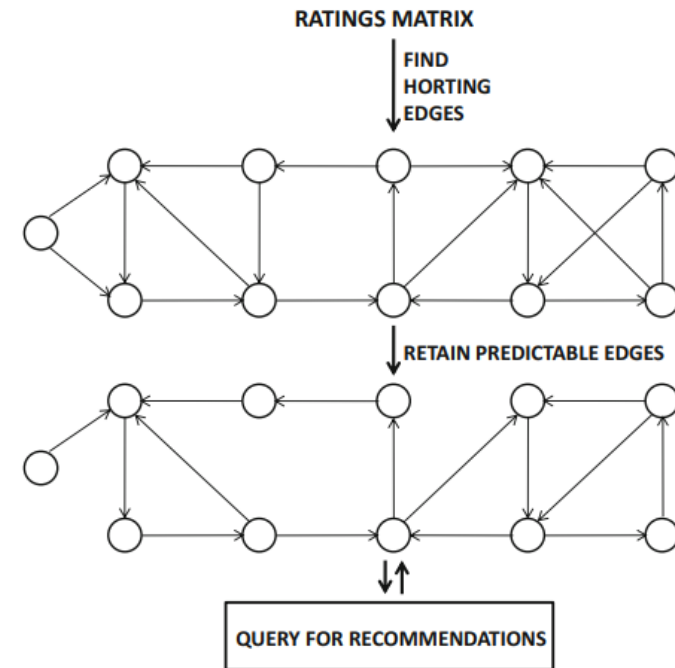


Figure 2.4: The user-user predictability approach