

추천 시스템

허석진

추천 시스템 (Recommender System)

이상품을 본 고객이 많이 본 다른 상품

2/4 < >



[위닉스] 제습기 9종

189,000원



[생활공작소] 제습제 520ml x 2...

15,300원



선반용/옷걸이형/서랍용 제습제

7,900원



위닉스 제습제

5,900원



사계절 뽕송하게 뽕송한데이 제...

8,900원

- 상품 / 검색 결과 노출 정렬
- 영화 / 뉴스 기사 추천

추천 시스템 종류

- 콘텐츠 기반 추천

- 기존에 이용했던 상품과 유사한 것 추천
- 아이템의 특성에 기반

- 협업 필터링 (Collaborative Filtering)

- 사용자간(User-User) 유사도 기반
- 상품간(Item-Item) 유사도 기반

협업 필터링

- 사용자 기반

- 각 사용자에게 대해 상품에 관한 유사도가 높은 다른 사용자 사용
- “나와 비슷한 사람이 선호하는 상품 추천”
- 사용자별 상품 선호도 정보 전체가 모형에 포함

- 상품 기반

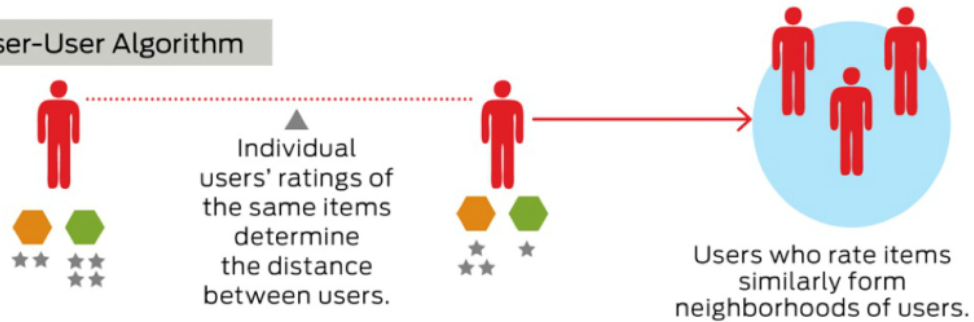
- 각 상품에 대해 사용자에게 관한 유사도가 높은 다른 상품 사용
- “어떤 상품에 관심있는 사람들이 선호하는 다른 상품 추천”
- 모형의 핵심은 상품간 유사도

협업 필터링

Item-Item Algorithm



User-User Algorithm



[출처 : [Deconstructing Recommender Systems](#) by JOSEPH A. KONSTAN, JOHN RIEDL / OCTOBER 2012]

사용자 기반 (User Based) 협업 필터링

- 사용자 u 의 상품에 i 에 대한 선호도 예측
 1. 사용자별 선호도를 정규화(normalize)
 2. 사용자 u 와 훈련 데이터 내의 사용자간 유사도 계산
 3. 유사도를 가중 평균

$$\frac{\sum_v r_{v,i} \cdot s_{u,v}}{\sum_v s_{u,v}}$$

$r_{v,i}$: 사용자 v 의 상품 i 에 대한 선호도

$s_{u,v}$: 사용자 u 와 사용자 v 사이의 유사도

상품 기반 (Item Based) 협업 필터링

- 사용자 u 의 상품에 i 에 대한 선호도 예측
 1. 사용자별 선호도를 정규화(normalize)
 2. 상품간 유사도 계산 (사전에 계산 가능)
 3. 유사도를 가중 평균

$$\frac{\sum_j r_{u,j} \cdot s_{i,j}}{\sum_j s_{i,j}}$$

$r_{u,j}$: 사용자 u 의 상품 j 에 대한 선호도

$s_{i,j}$: 상품 i 와 상품 j 사이의 유사도

유사도 (similarity)

- Euclidean Distance

- 기하학적인 거리: $\sqrt{\sum_i (x_i - y_i)^2}$

- Pearson Correlation

- 공통된 항목에 대한 상관계수: $\frac{cov(x,y)}{\sigma_x \sigma_y}$

- Cosine

- 벡터들의 cosine (빠진 항목은 0으로 간주): $\frac{x \cdot y}{|x||y|}$

- Jaccard

- 전체에서 공통된 항목의 비중: $\frac{|X \cap Y|}{|X \cup Y|}$

모형 평가 (Evaluation) - 정확도

- MAE (mean absolute error)
 - 기준값과의 차이들의 평균
- MSE (mean squared error)
 - 기준값과의 차이들의 제곱 평균
- RMSE (root mean squared error)
 - \sqrt{MSE}

모형 평가 (Evaluation) – 정밀도와 재현율

- 정밀도(precision)

- 긍정 예측(추천) 중에서 실제 관측된(높은 선호도) 비율

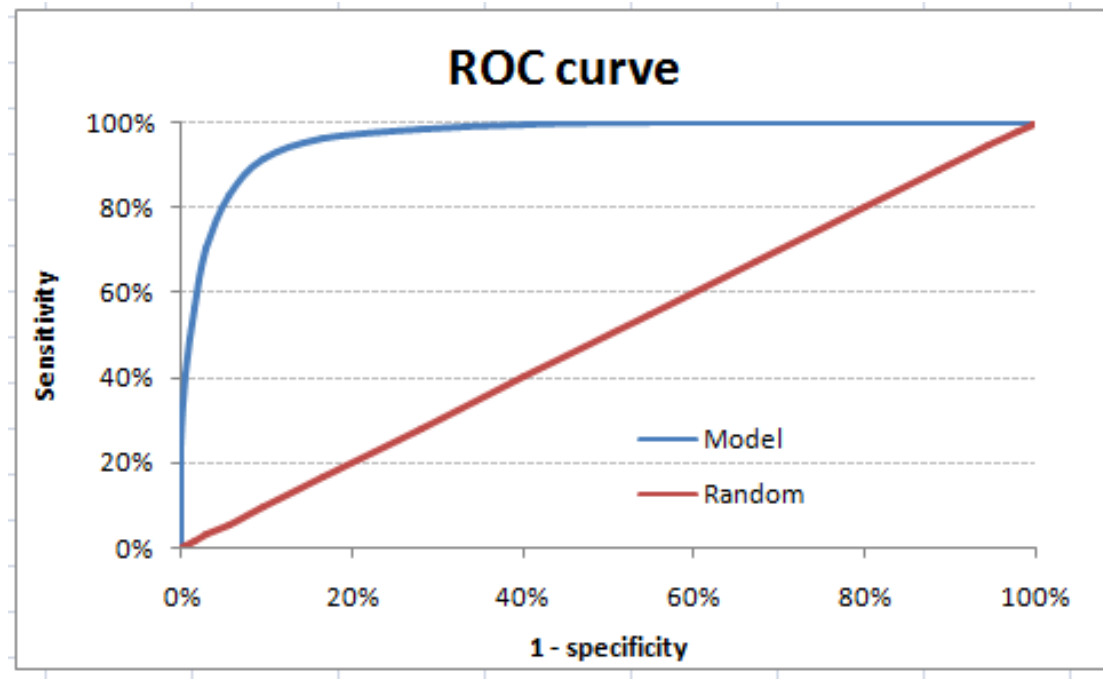
- > 잘못된 예측이 얼마나 적은지

- 재현율(recall)

- 실제 관측된(높은 선호도) 경우 긍정으로 예측(추천)한 비율

- > 빼먹은 예측이 얼마나 적은지

모형 평가 (Evaluation) – ROC



- 가로축: 실제 관측되지 않은(낮은 선호도) 경우 긍정으로 예측(추천)한 비율
- 세로축: 실제 관측된(높은 선호도) 경우 긍정으로 예측(추천)한 비율
- ROC 아래 영역의 넓이(AUC)를 모형 평가 지표로 사용

모형 평가 (Evaluation) – 순서

- Spearman's rank correlation
 - 상위 몇 개의 상품에 대해
 - 추천된 상품의 순서와 실제 선호도에 따른 순서 계산
 - 상관계수 산출
- > 추천된 순서와 실제 선호도의 순서가 얼마나 비슷한지

모형 평가 (Evaluation)

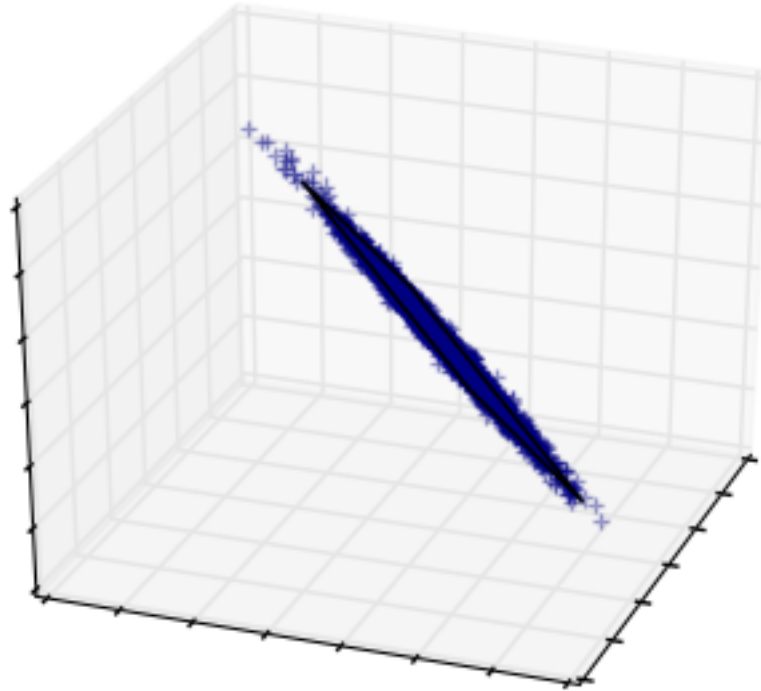
- 이슈

- 기준이 되는 사용자들의 선호도 필요
- 모든 사용자의 선호도를 사용하는 것은 현실적으로 불가능
- 실제로 모든 사용자의 선호도가 있다면 추천 시스템이 필요?

- 차선책

- 훈련-테스트 관계처럼 일부 가용한 사용자 선호도로 모형 평가
- 실질적으로 의미있는 방법은 사용자 반응에 기반한 효용 측정

차원 축소



- 3차원 공간의 점들
- 실제로는 1차원 데이터에 가까움

차원 축소

- 이슈

- 추천을 위한 유사도 계산 시 사용자-상품 행렬의 크기가 문제
- 사용자-상품 행렬에 포함된 노이즈에 의한 과적합

- 개선

- 원래의 사용자-상품 행렬의 정보를 대부분 유지하면서 행렬과 노이즈의 크기를 줄이는 방안 필요

- 특이값

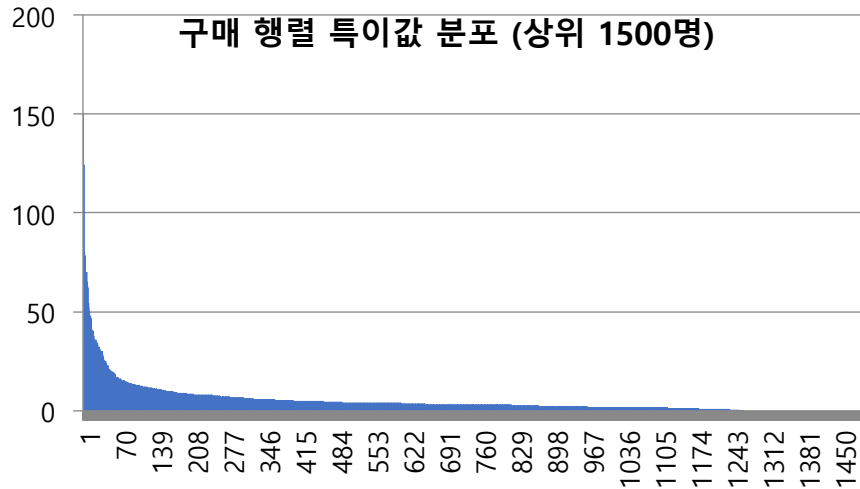
- 0이 아닌 특이값들이 실질적인 차원을 나타냄

특이값 분해

The diagram illustrates the Singular Value Decomposition (SVD) of a matrix A . It shows the equation $A = U \Sigma V^T$ with the dimensions of each matrix specified below them.

- A is an $m \times n$ matrix.
- U is an $m \times m$ matrix.
- Σ is an $m \times n$ matrix, represented by a diagonal of yellow squares. It is flanked by yellow 0 symbols indicating zero padding.
- V^T is an $n \times n$ matrix.

특이값 분해



• User-Item 행렬 축소

- 협업 필터링에 사용되는 User-Item 행렬의 특이값 분해 분석
- 상위 구매 사용자 집중으로 과적합 방지와 계산량 절약

