Collaborative Filtering in RS

Collaborative filtering - topics

- Introduction to collaborative filtering and how it works in RS
- Matrix Factorization Based approach using SVD
- Performing SVD on different dataset
- Hands on exercise on collaborating filtering models

Introduction to collaborative filtering

- Collaborative filtering, also referred to as social filtering, filters information by using the recommendations of other people.

- For each user, recommender systems recommend items based on how similar users liked the item.

For eg:, A person who wants to see a movie for example, might ask for recommendations from friends. The recommendations of some friends who have similar interests are trusted more than recommendations from others. This information is used in the decision on which movie to see.

Introduction to collaborative filtering

Types of CF

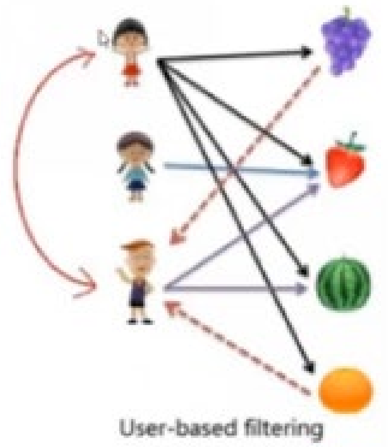
- Item based CF : Compute similarity between items
- User based CF : Compute similarity between users

User based CF

- Identify users who have a similar taste of products.

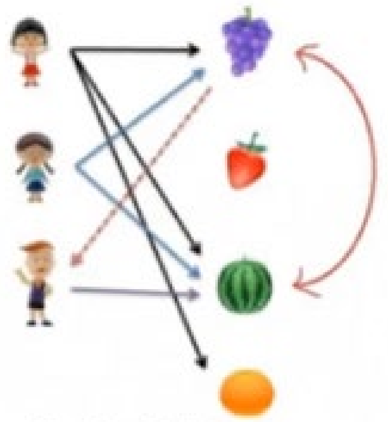
- Similarity of the users is based upon their purchasing behaviour.

- Memory-based: the rating matrix is directly used to find neighbours / make predictions.



Item based CF

- Users are recommended based on the item they bought.
- Similarity is based upon the co-occurence of purchases:
  - Similarity between items (and not users) is considered to make predictions



Item-based filtering

How to compute similarity

- Cosine similarity

$$CosSim(x,y) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}} = \frac{\langle x, y\rangle}{||x||\ ||y||}$$

- Pearson similarity

$$\begin{aligned} Corr(x,y) &= \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2}\sqrt{\sum (y_i - \bar{y})^2}} \\ &= \frac{\langle x - \bar{x},\ y - \bar{y}\rangle}{||x - \bar{x}||\ ||y - \bar{y}||} \\ &= CosSim(x - \bar{x}, y - \bar{y}) \end{aligned}$$

User based CF Vs. Item based CF

IBCF is more efficient than UBCF

- Typical applications involve far more Users than items. Hence Similarity matrix for IBCF is more compact than UBCF.
- Similarity estimates between items is more likely to converge over time than similarity between users.
- However, the IBCF recommendations tend to be more conservative than UBCF.

Strengths and issues of CF

Strengths

- Content- agnostic
- Does not require items or users to be related with the content information.
- Recommendations are very personalised in this case

Issues

- Cold start problem
- Sparsity
- Popularity bias
- Scalability

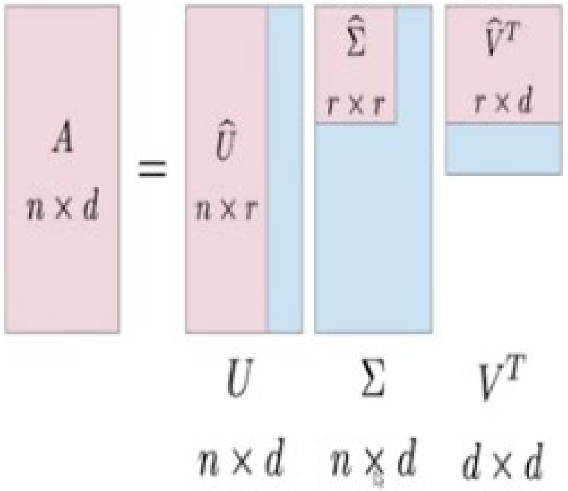
Matrix Factorization Based approach using SVD

What is SVD?

SVD can be used to decompose any given matrix, M into a product of 3 matrices as follows:

$$M = U \times \Sigma \times V^T$$

Where U and V are called left and right singular vectors.



Hybridization Methods

Methods of Hybridization

- Weighted - Recommendations from each system is weighted to calculate final recommendation.
- Switching- System switches between different recommendation model.
- Mixed - Recommendations from different recommenders are presented together.

Hybrid recommender systems

Multiple recommender systems are combined to improve recommendations.

- Although any type of recommender systems can be combined a common approach in industry is to combine content based approaches and collaborative filtering approaches.
- Content based models can be used to solve the Cold start and Grey sheep problems in Collaborative filtering.

Case Study

Objective:

- Generate a top-n list of restaurants on consumer preference - Restaurant & Consumer data using the data in rating.final.csv

Abstract:

- The dataset was obtained from a recommender system prototype. The task was to generate a top-n list of restaurants according to the consumer preferences.
- The collaborative filter technique used only one file i.e., rating_final.csv that comprises the user, item and rating attributes

Case Study Contd.

Attribute Information:

- UserID – User id, object type
- placeID – place id, int type
- Rating – overall rating, int type
- Food_rating – food rating, int type
- Service_rating – service rating, int type

HappyLearning

Questions?