MATRIX FACTORIZATION TECHNIQUES FOR RECOMMENDER SYSTEMS

Yehuda Koren, Yahoo Research
Robert Bell and Chris Volinsky, AT&T Labs - Research
Recommender System Strategies

Content Filtering Approach

Collaborative Filtering Approach

Neighborhood Methods

Latent Factor Methods

Matrix Factorization Method
Content Filtering Approach

Creates a profile for each user or product to characterize its nature

- Eg. Movie profile include attributes regarding its genre, the participating actors, its box office popularity...

- Eg. User profile might include demographic information or answers provided on a suitable questionnaire
Collaborative Filtering Approach

Relies only on past user behavior (eg. Previous transactions or product ratings)

Basic assumption and idea

- Users give ratings to catalog items
- Customers who had similar tastes in the past, will have similar tastes in the future
User Oriented Neighborhood Method
Recommender System Strategies

- Content Filtering Approach
- Collaborative Filtering Approach
  - Neighborhood Methods
  - Latent Factor Methods
    - Matrix Factorization Method
Latent Factor Method

The diagram illustrates the positioning of various movies on a four-dimensional space defined by the factors: Serious vs. Escapist and Geared toward females vs. Geared toward males.

- Serious: Amadeus, Braveheart, Ocean's 11, The Lion King, Independence Day, Dumb and Dumber
- Escapist: The Color Purple, Sense and Sensibility, The Princess Diaries, Gus
- Geared toward females: The Color Purple, Sense and Sensibility, The Princess Diaries
- Geared toward males: Amadeus, Braveheart, Ocean's 11, The Lion King, Dumb and Dumber
Matrix Factorization Methods

Characteristic

○ Characterizes both items and users by vectors of factors inferred from item rating patterns

Rely on matrix types of input data

○ One dimension representing user
○ The other representing items

Two data types

○ High-quality explicit feedback
  ▪ Includes explicit input by users regarding their interest in products
  ▪ We refer to explicit user feedback as ratings
  ▪ Usually sparse matrix, since any single user is likely to have rated only a small percentage of possible items
Implicit feedback

Which indirectly reflects option by observing user behavior

- Purchase history, browsing history, search patterns, mouse movements
  - Usually denotes the presence or absence of an event
  - Typically represented by a densely filled matrix
Difficulties

- High portion of missing values caused by sparseness in the user-item rating matrix
1. Adding Biases

Typical collaborative filtering data exhibits large systematic tendencies for some users to give higher ratings than others.

And for some items to receive higher ratings than others.

Some products are widely perceived as better (or worse) than others.

It's unwise to explain the full rating value in this form.

We should identify the user and item bias.
Matrix Factorization Methods

Adding Biases

For example:
- Suppose we want to estimate user John’s rating of the movie Titanic
- And the average rating over all movies is 3.7 stars
- Titanic is better than an average movie, so it tends to be rated 0.5 starts above the average movie
- John is a critical user, who tends to rate 0.3 stars lower than the average
- Thus, the estimate for Titanic’s rating by John would be (3.7 + 0.5 - 0.3)
2. Additional Input Sources

A system must deal with cold start problem, wherein many users supply very few ratings

- We need to incorporate additional sources to relieve this problem
- Using implicit feedback to gain insight into user preferences
- Eg: Demographics, such as gender, age group, zip code, income level.
3. Temporal Dynamics

Time-drifting nature

- So far, the presented models have been static
- In reality, product perception and popularity constantly change as new selections emerge
- Similarly, customer’s inclinations evolve, leading them to redefine their taste
- The system should account for the temporal effects
Matrix Factorization Methods

Temporal Dynamics

Multiple sources of temporal dynamics

- **Item-side effects:**
  - Product perception and popularity are constantly changing
  - Seasonal patterns influence items’ popularity

- **User-side effects:**
  - Customers ever redefine their taste
  - Drifting rating scale
  - Change of rater within household
Matrix Factorization Methods

4. Inputs With Varying Confidence Levels

- Are all observed ratings deserve the same weight or confidence?
  - For example, massive advertising might influence votes for certain items, which do not aptly reflect longer-term characteristics.

- The user might be unaware of the existence of the item, or unable to consume it due to its price.
  - Consuming an item can also be the result of factors different from preferring it:
    - A user may watch a TV show just because he/she is staying on the channel of the previously watched show.
    - A consumer may buy an item as a gift for someone else, despite not liking the item for himself.
Conclusion

Matrix Factorization techniques have become a dominant methodology within collaborative filtering recommenders.

At the same time, they offer a compact memory-efficient model that systems can learn relatively easily.

What makes these techniques even more convenient is that models can integrate naturally many crucial aspects of the data, such as multiple forms of feedback, temporal dynamics, and confidence levels.