Overview: Collaborative Filtering

Collaborative filtering is a technique that can filter out items that a user might like on the basis of reactions by similar users.

Collaborative filtering analyzes relationships between users and interdependencies among products to identify new user-item associations.

The algorithm uses “user behavior” for recommending items.

It's based on the idea that people who agreed in their evaluation of certain items are likely to agree again in the future.

Approach

- Neighborhood Methods
- Latent Factor Models

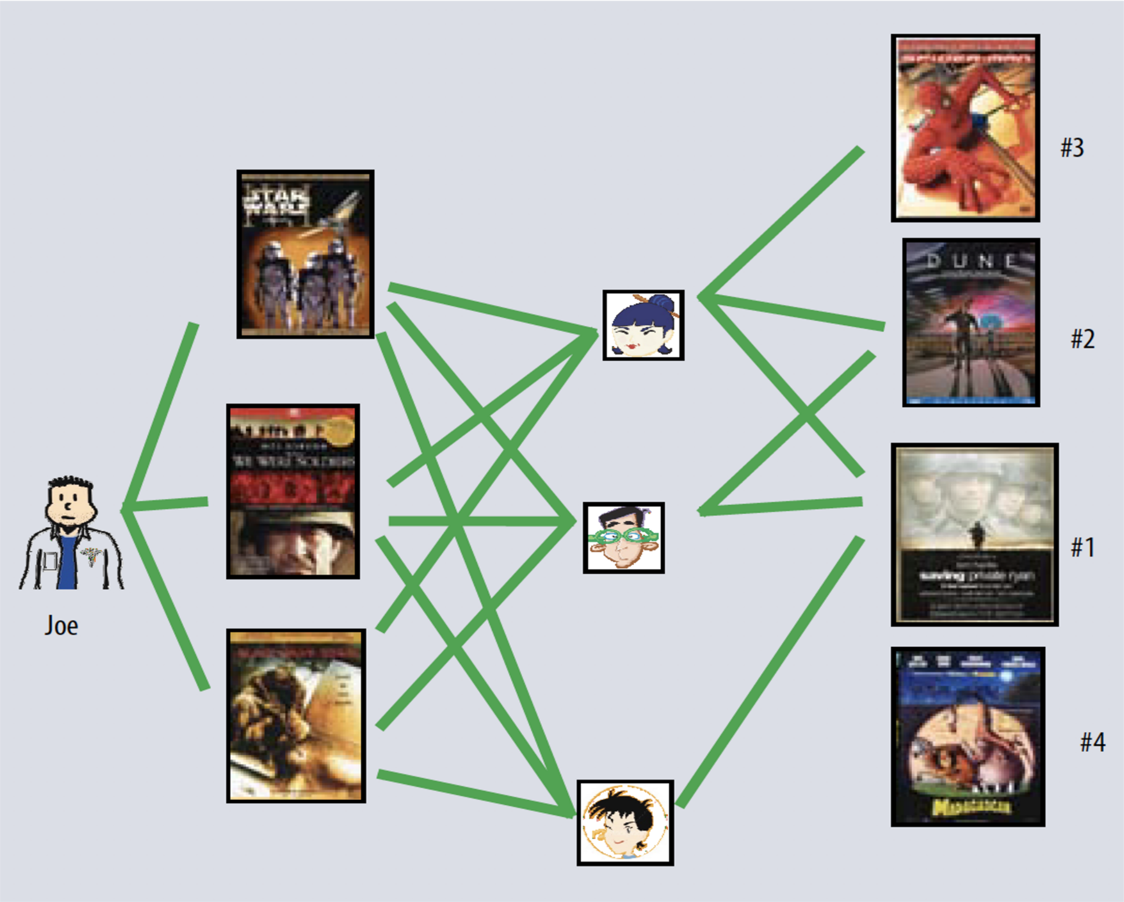


Figure 1. The user-oriented neighborhood method. Joe likes the three movies on the left. To make a prediction for him, the system finds similar users who also liked those movies, and then determines which other movies they liked. In this case, all three liked *Saving Private Ryan*, so that is the first recommendation. Two of them liked *Dune*, so that is next, and so on.

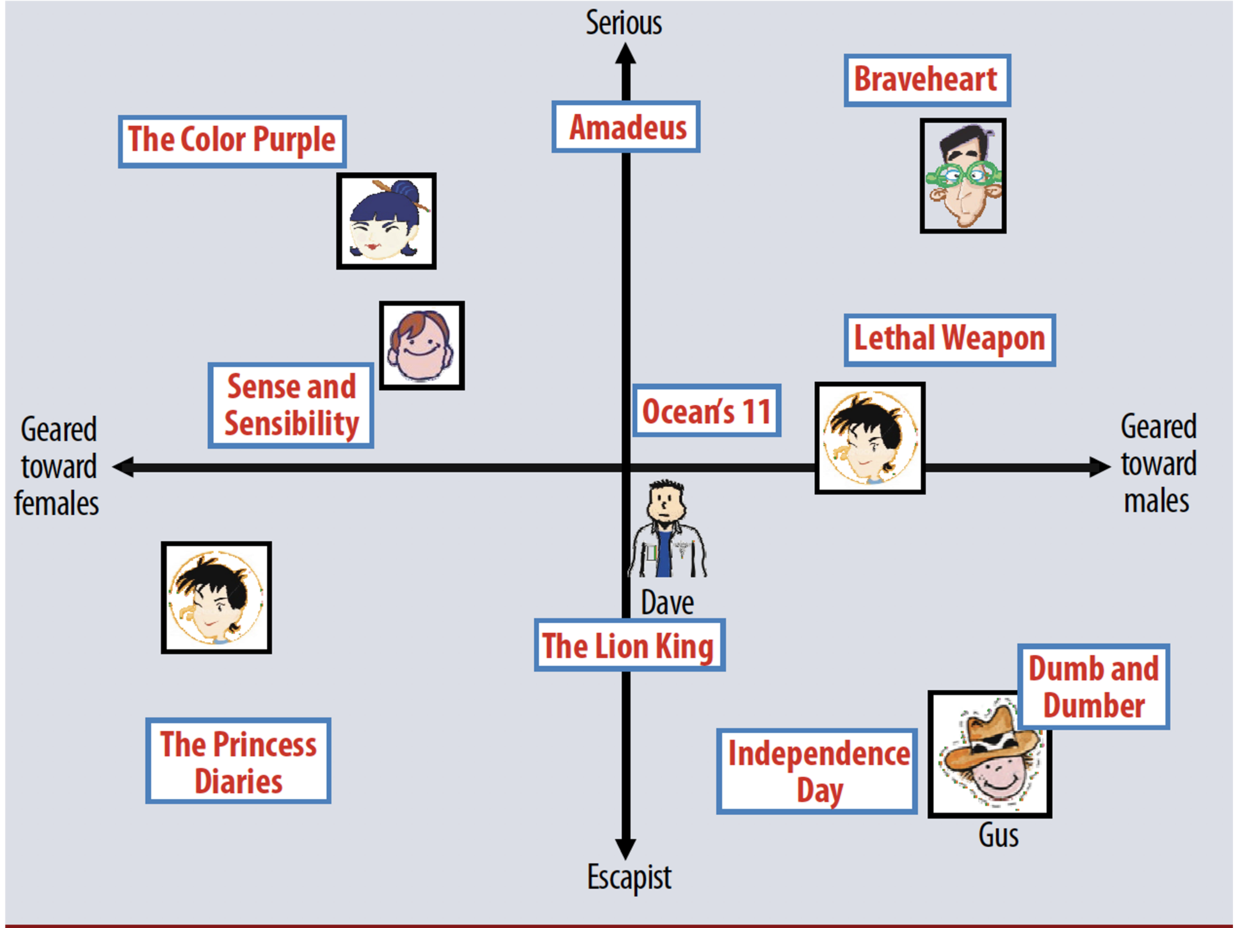


Figure 2. A simplified illustration of the latent factor approach, which characterizes both users and movies using two axes—male versus female and serious versus escapist.

Advantage & Disadvantage

- **Advantage**

Does not need to access side information: not required to understand item content. The content of the items does not necessarily tell the whole story, such as movie type/genre, and so on.

- **Disadvantage**

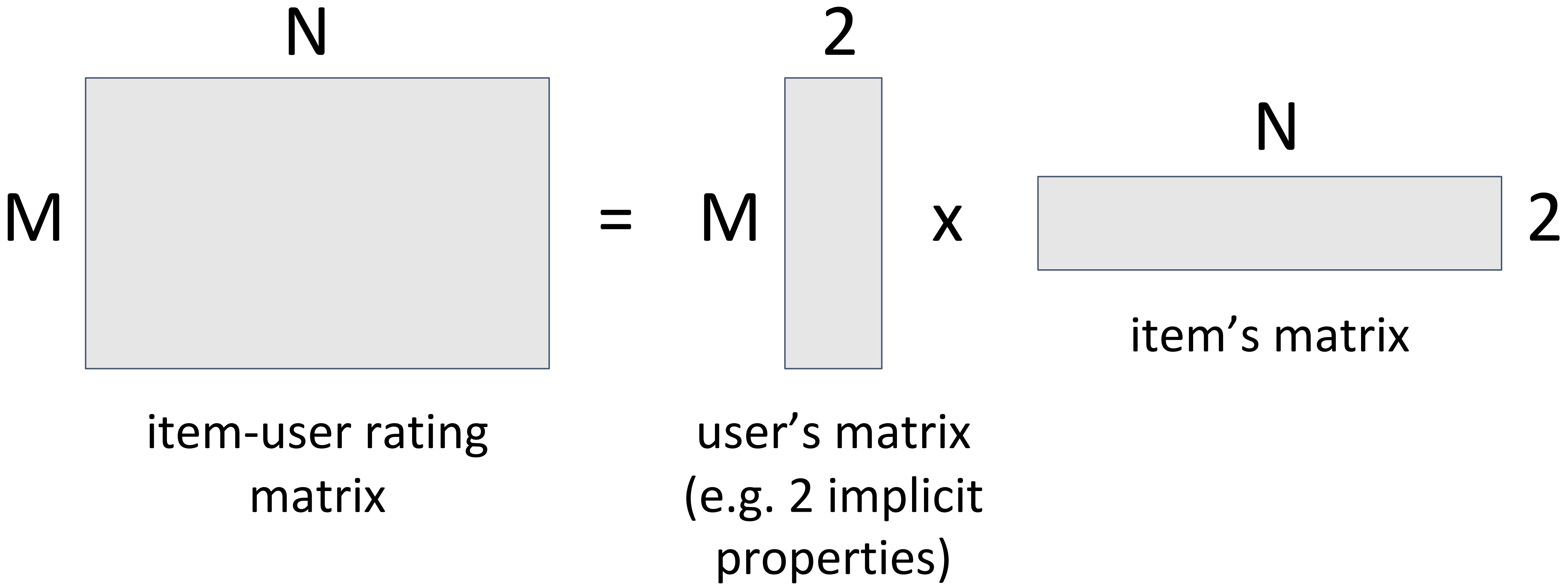
Cold start problem: inability to address the system’s new products and users. If an item is not seen during the training, the system cannot create an embedding for it and cannot query the model with this item.

Matrix Factorization Methods

Matrix factorization characterizes both items and users by vectors of factors inferred from item rating patterns. High correspondence between item and user factors leads to a recommendation.

Matrix factorization models map both users and items to a joint latent factor space of dimensionality f, such that user-item interactions are modeled as inner products in that space.

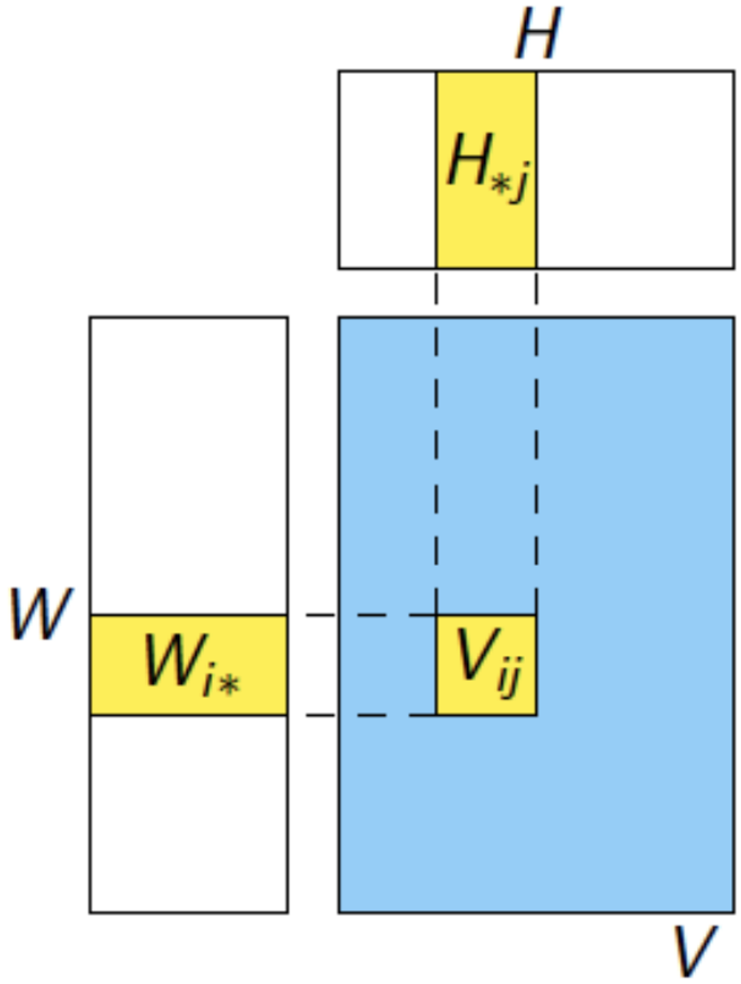
Matrix Factorization Illustration



Matrix Factorization Illustration

Users vectors: $(W_{u*})^T \in \mathbb{R}^r$

Items vectors: $H_{*i} \in \mathbb{R}^r$



Matrix Factorization Illustration

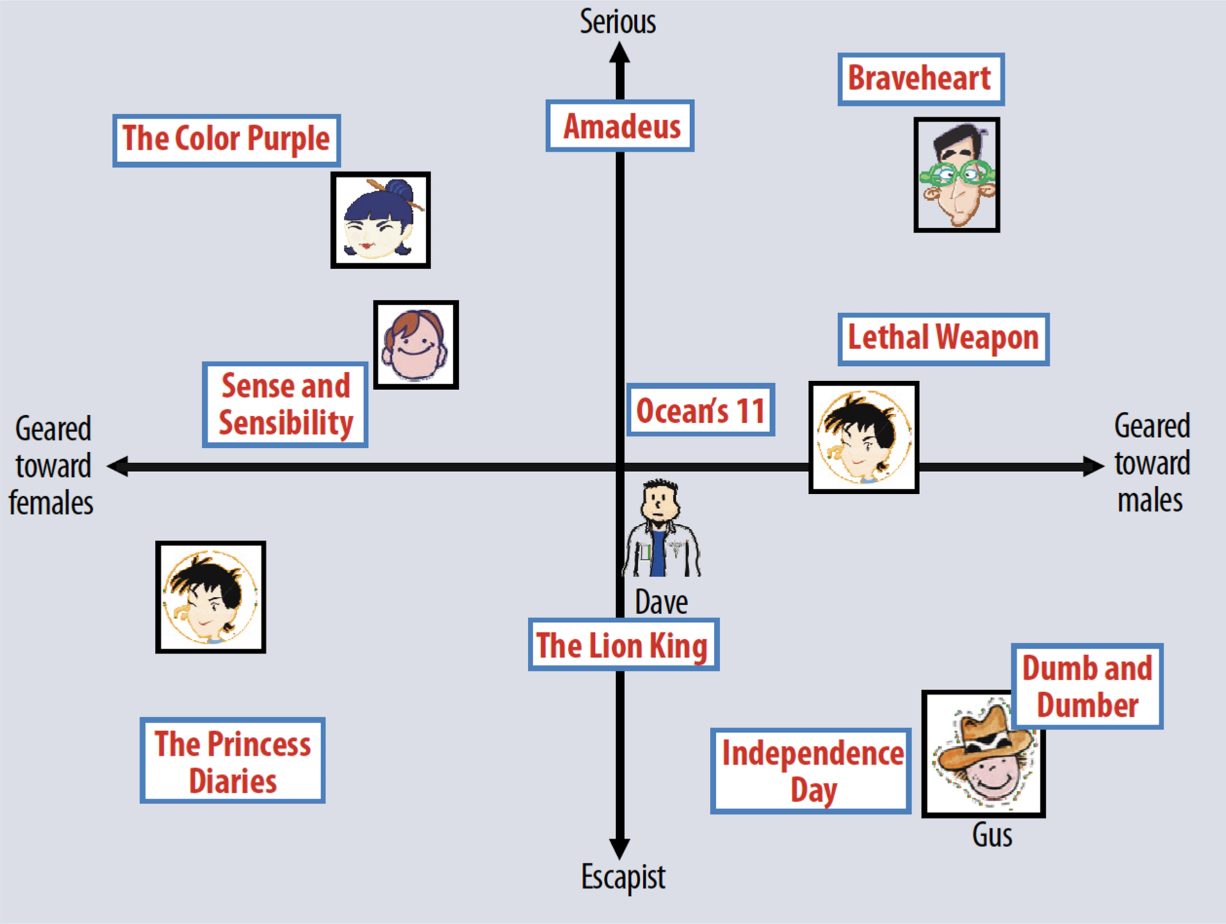


Figure 2. A simplified illustration of the latent factor approach, which characterizes both users and movies using two axes—male versus female and serious versus escapist.

- Serious vs. Escapist
- Males vs. Females

- Users classified into 4 quadrants

Method

- Each item i is associated with a vector q_i and each user u is associated with a vector p_u
- For a given item i, the elements of q_i measure the extent to which the item possesses those factors. For a given user u, the elements of p_u measure the extent of interest the user has in items that are high on the corresponding factors.
- Approximates user u's rating of item i: $\hat{r}_{ui} = q_i^T p_u$.