COLLABORATIVE FILTERING

NAME:NIRAJAN PRAJAPATI

ROLL NO:KCE074BCT024

Introduction

- Collaborative filtering is a technique that can filter out items that a user might like on the basis of reactions by similar users.
- It works by searching a large group of people and finding a smaller set of users with tastes similar to a particular user. It looks at the items they like and combines them to create a ranked list of suggestions.

Steps involve in collaborative filtering

- The first step is to find similar users or items.
- The second step is to predict the ratings of the items that are not yet rated by a user.

Types of filtering

- User-user filtering
  - For a user **U**, with a set of similar users determined based on rating vectors consisting of given item ratings, the rating for an item **I**, which hasn't been rated, is found by picking out N users from the similarity list who have rated the item **I** and calculating the rating based on these N ratings.
- Item-item filtering
  - For an item **I**, with a set of similar items determined based on rating vectors consisting of received user ratings, the rating by a user **U**, who hasn't rated it, is found by picking out N items from the similarity list that have been rated by **U** and calculating the rating based on these N ratings.

- Item-item filtering is effective because usually, the average rating received by an item doesn't change as quickly as the average rating given by a user to different items.

Types of algorithm

- Memory based algorithm
- Model based algorithm

Memory based Algorithm

- To find the rating **R** that a user **U** would give to an item **I**, the approach includes:
  - Finding users similar to **U** who have rated the item **I**
  - Calculating the rating **R** based the ratings of users found in the previous step.

- The data includes four users **A**, **B**, **C**, and **D**, who have rated two movies. The ratings are stored in lists, and each list contains two numbers indicating the rating of each movie:
  - Ratings by **A** are [1.0, 2.0].
  - Ratings by **B** are [2.0, 4.0].
  - Ratings by **C** are [2.5, 4.0].
  - Ratings by **D** are [4.5, 5.0].

Memory based

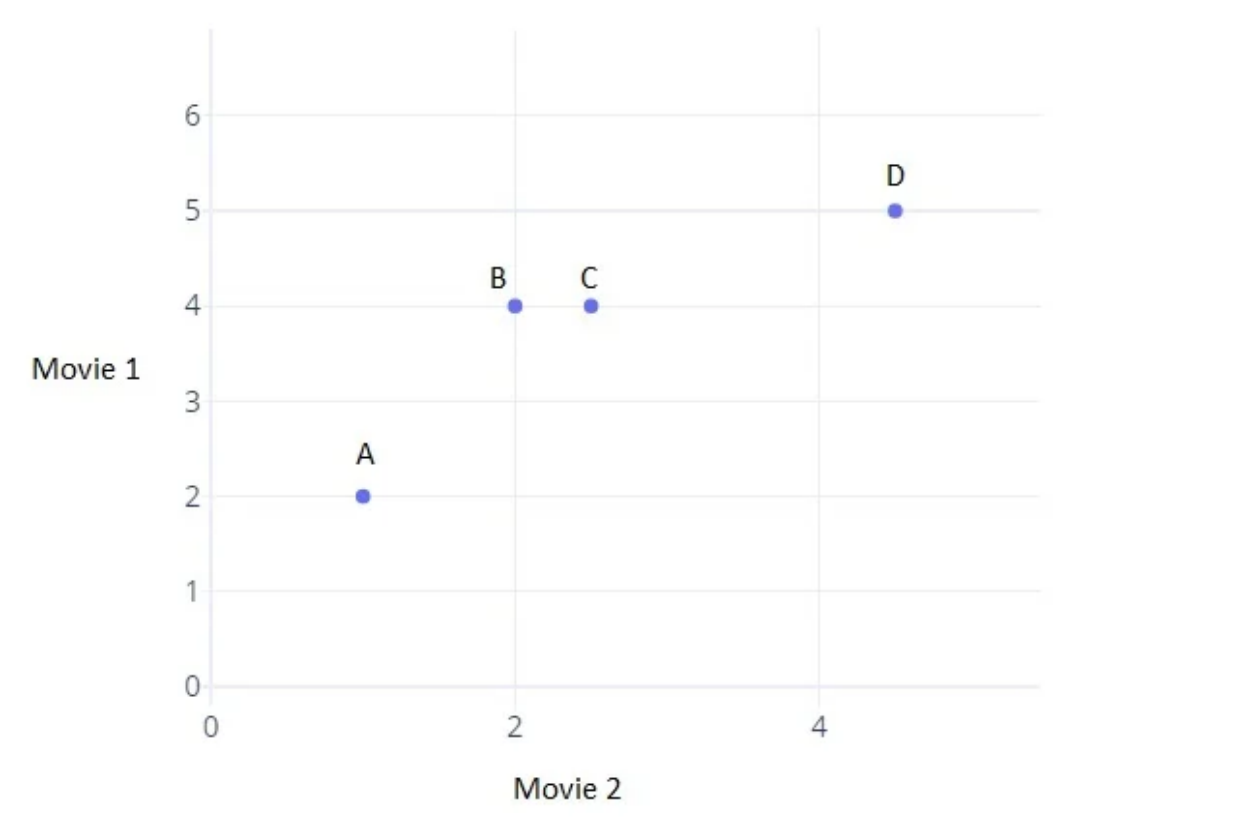


Fig:using distance

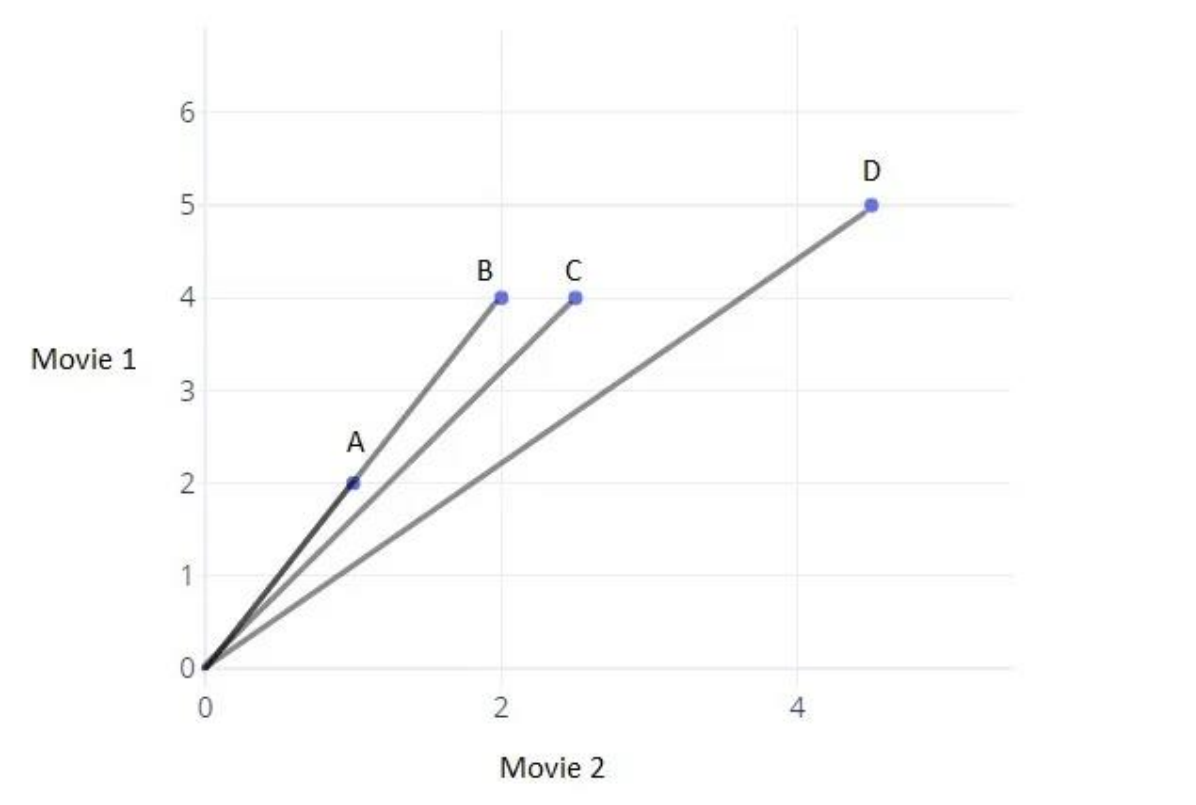


Fig:using cosine

Model based

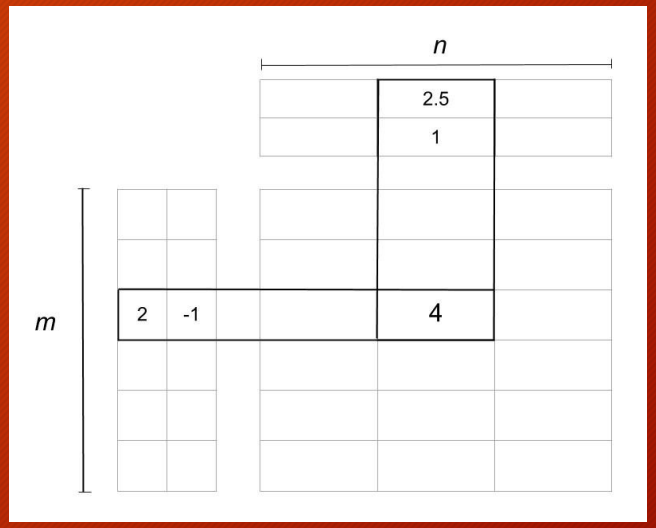
- It involves step to reduce or compress the large but sparse user-item matrix
- It uses user-item matrix to predict rating.

Matrix factorization algorithm

- Matrix factorization mean breaking down the bigger matrix into smaller matrix.
- For example: matrix a x b can be represented as a x m and m x b

Matrix factorization

- The reduced matrices actually represent the users and items individually. The **m** rows in the first matrix represent the **m** users, and the **p** columns tell you about the features or characteristics of the users. The same goes for the item matrix with **n** items and **p**



Matrix factorixation

Fig: without dimension reduction

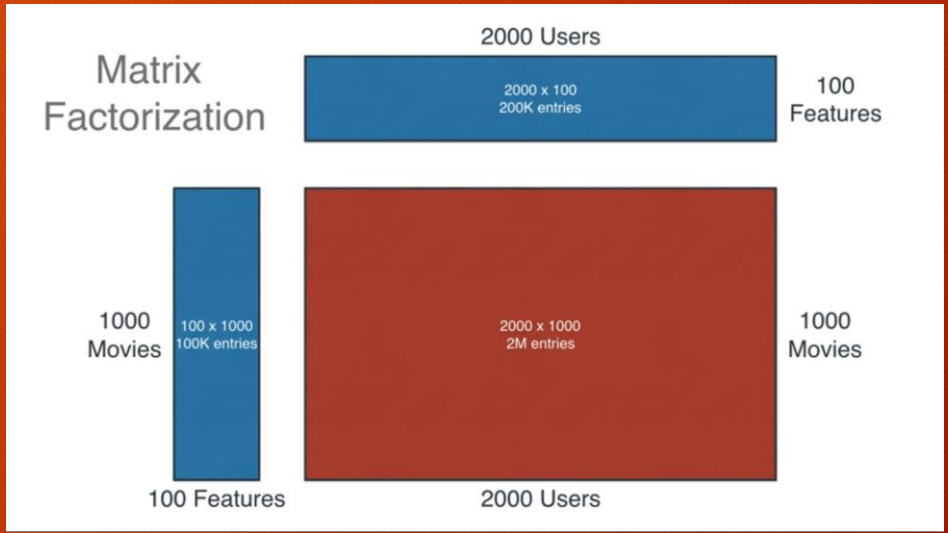
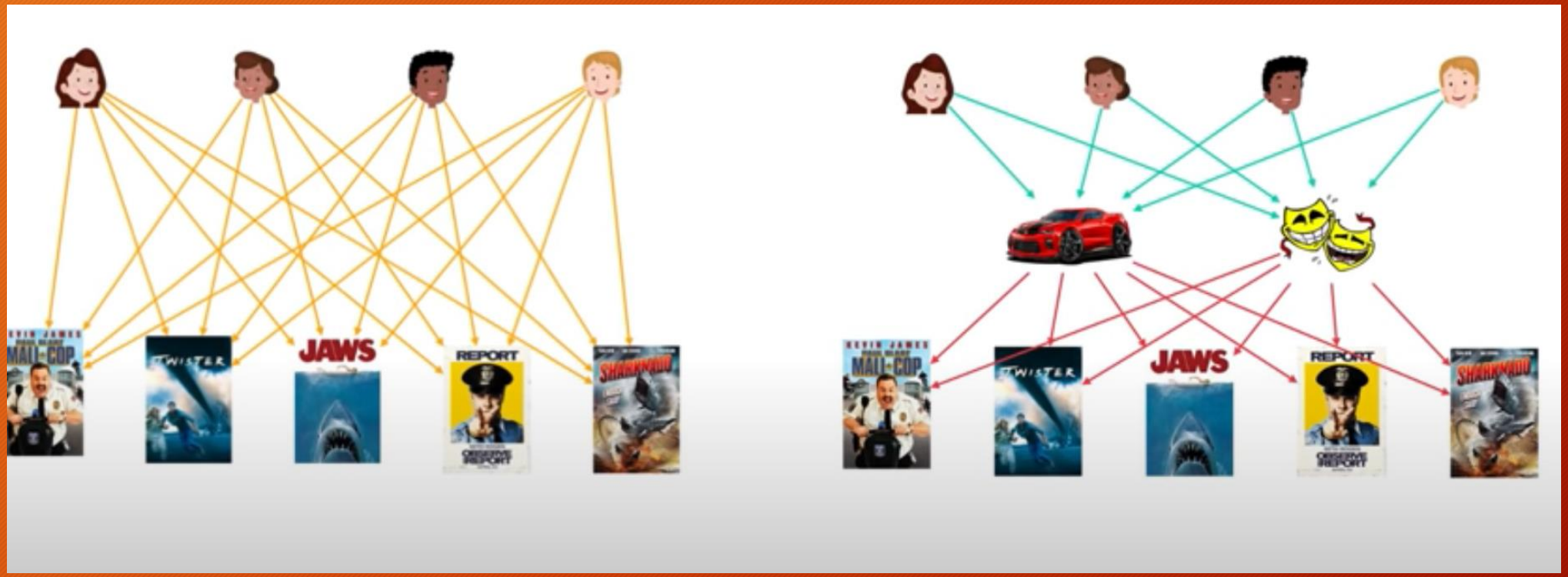


Fig: with dimension reduction

Matrix factorization



Problems in collaborative filtering

- Collaborative filtering can lead to some problems like cold start for new items that are added to the list. Until someone rates them, they don’t get recommended.

Error calculation

- RMSE
- Mae

Thank you

ANY QUESTION?