Movie Recommendation System Report

1. Introduction

When we want to watch a movie, it's often hard to pick one we like. That’s why recommendation systems are important. In this project, I tried three different ways to recommend movies:

a.User-Based Filtering: Find users who like the same things as you and recommend what they like.

b.Item-Based Filtering: Find movies that are similar to the ones you've already liked.

c.Random Walk: Imagine walking randomly from your favorite movie to other movies on a graph — the ones you reach the most are recommended.

2. Dataset Description

I used the MovieLens 100K dataset. It includes:
943 users
1,682 movies
100,000 ratings

I first turned the data into a rating matrix, a big table where each row is a user and each column is a movie, showing how they rated it. Then, I built a graph where users and movies are nodes, and an edge connects them if the user rated the movie.

3. Methodology

User-Based: calculated how similar two users are. If someone is similar to user, the movies they like might be good for this user too.

Item-Based: looked at which movies are often liked by the same person. If a user like Movie A, then recommend other movies that are similar to A.

Random Walk: used a graph structure. Starting from movies user’ve liked, “walked” randomly through the graph. Movies that were visited more often are considered more related to this user.

4. Implementation Details

Step 1: Prepare the Rating Data

First, I used the pandas library to read the MovieLens 100K dataset. This dataset tells us how many stars each user gave to each movie. Then, I created a rating matrix where the rows are users and the columns are movies. This matrix helps us find which users or movies are similar to each other.

Step 2: Build the Graph (Adjacency List)
To use the Random Walk method, I needed to turn the data into a graph. In this graph:
There are two types of nodes: users and movies.
There is an edge between a user and a movie if that user has watched and rated the movie.
So the graph is like a map that connects people and the movies they watched.

Step 3: Perform the Random Walk
I started the random walk from the movies that the user liked the most. Then:
Randomly picked a connected node,could be another user or another movie,
Took one step and recorded the movie I landed on,
Repeated this walking process many times.
At the end, counted how many times each movie was visited. The more a movie was visited, the more likely it is interesting to the user, and used this to make recommendations.

Step 4: Filter Out Watched Movies
Before showing the final recommendations, I removed all the movies the user had already watched. Then only kept the movies they haven't seen, sorted by how many times they were visited during the random walk. These became the final recommended movies.

5. Results and Evaluation

From the results, these three methods have their benifit in different situation.
User-Based: Easy to understand and use, doesn't work well for new users.
Item-Based: Stable results, doesn't care much about user's taste.
Random Walk: Can find new connections, but slower and needs a graph to be ready.

6. Conclusion

In this project, I learned about three different recommendation methods, each with its own strengths:
User-based is the easiest to start with
Item-based gives the most stable results
Random walk is creative and powerful

In the future, we could combine these methods together and add more features like movie genre, actors, etc., to improve accuracy. These methods are used in real apps like Netflix.