Name: Sravanthi Kapu

Project Title: Movie recommendation System

Project Design

Movie recommendation system: The idea is to build a recommendation system which recommends movies to a user, based on other user's previous ratings. Implementation to find the movies which user might be interested will be done by using collaborative filtering.

In a movie recommendation system, many users rating's for many movies is analysed. By comparing the users who are having similar preferences, movies can be recommended to a user based on the ratings and genres liked by the user. The approach called Collaborative filtering is used in order to achieve this.

Dataset: Movielens data set is used. A dataset with all the data related to movies, rankings, genre, user id and other fields is used.

https://grouplens.org/datasets/movielens/

Implementation of the Project:

Step1: The data set movies and rating files are used to load the dataset and then build a ratings matrix with users as rows, movies as columns, and ratings as the elements of the matrix.

Step2: Now we will predict two matrices: The k × n user matrix X, and the k × m movie matrix Y with random values.

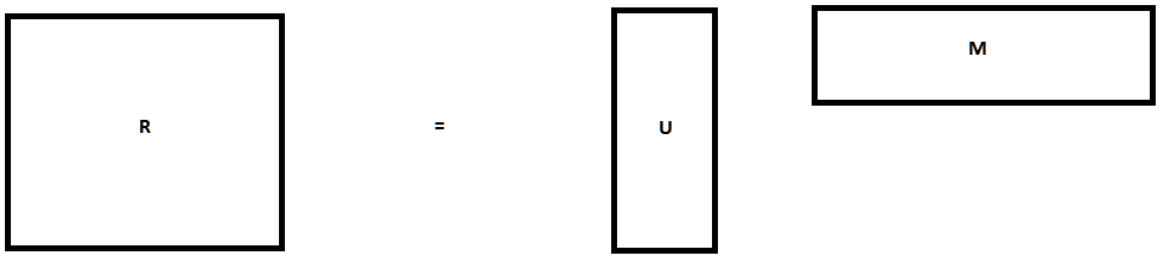
Step3: Using the Alternating least squares method, we will perform matrix factorization in order to obtain ratings matrix. This step is repeated until the RMSE value is low; on an average of 18 to 25 iterations can be done in order to reach least RMSE value.

The alternating least squares and Matrix factorization are explained below.

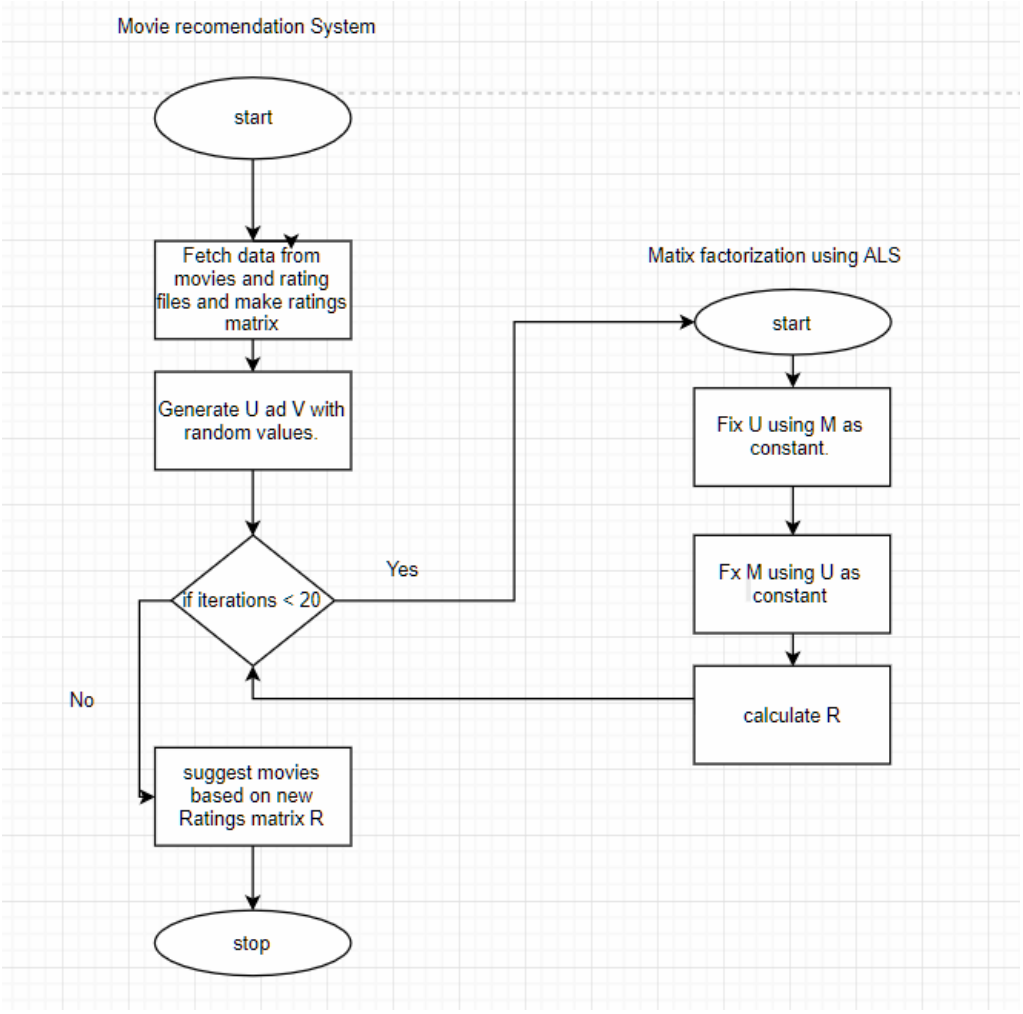
Step4: Now the new ratings matrix is calculated. Now among those missing values which are filled, the movies with highest rating can be suggested to user as recommendations.

Matrix Factorization:

We are going to approximate our given, sparse, R matrix as the product of two smaller matrices. You can consider them a user feature matrix and movie feature matrix. This approximation is also going to smooth out the zeros and in the process give us our projected ratings.



High level Description:



Algorithms Used:

In order to predict the suggestions for the user, alternate least squares algorithm is used.

We will implement alternating least squares approach with regularization. By doing so, we first estimate U using M and estimate M by using U. After enough number of iterations, we are aiming to reach a convergence point where either the matrices U and M are no longer changing or the change is quite small.

Output: The final obtained matrix, all the missing values will be calculated by ALS matrix factorization using latent matrices and is used to recommend movies to user.