**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**

**I Semester, 2019-20**

# Introduction:

This project aims to build analyse the results of the following recommender systems which are used to predict unknown ratings of movies.

1. Collaborative Filtering.
2. Collaborative Filtering using Baseline approach.
3. Singular Value decomposition.
4. SVD with 90% energy.
5. CUR.
6. CUR with 90% energy.
7. Latent Factor Model

# Architecture:

Programming Language: Python

## Dataset:

Movie rating dataset is used in this assignment consisting of 6040 users rating of 3883 movies

Movie lens Dataset: [https://grouplens.org/datasets/movielens/](about:blank)

## Libraries used:

* Numpy
* Time
* Random

## Data Structures used:

* Np array
* Python list

## 

## **Working Description:**

Pre-processing:

The three dataset files users.dat, ratings.dat and movies.dat have been taken input as pandas dataframes. Using these dataframes, a new dataframe utility\_matrix has been built. It contains userIds as rows, movieIds as columns and their ratings at their respective positions in the dataframe.

This matrix uses the userId, movieId and their ratings present in ratings.dat file to fill the rating values in the utility\_matrix.

The data set is split into training and test set.training set is used to decompose into respective matrices and the reconstructed matrix is used to predict the values of the test set.

### Collaborative Filtering:

Collaborative filtering offers a way of predicting the value(to be given) of an entity for an item based on the value given by other similar entities to the corresponding item.

### Singular Value decomposition:

The goal of SVD is to decompose M as the product of three matrices Umxm, Smxn and VTnxm  and reduce the re construction error

The user ratings are predicteed and error is calculated

### CUR:

Aim of CUR decomposition is to reduce the density of SVD decomposition without having much loss of data. CUR decomposition aims to decompose the given matrix M into 3 matrices C, U, R, where, C is matrix comprising of randomly selected columns and R is matrix comprising of randomly selected rows

Number of rows selected is approximately equal to the rank of utility matrix

The original matrix is constructed by multiplying the C,U,R matrix

### Latent Factor Model:

This model uses matrix factorization techniques to find out latent or hidden factors of users and movies that determine the user ratings to movies. This is similar to singular value decomposition. SVD is usually more prone to overfitting and also requires factorization and this causes difficulties because of the sparseness of the user-movie matrix. It does not scale well with increasing data. This model avoids overfitting by regularization.

## **Results:**

|  |  |  |  |
| --- | --- | --- | --- |
| Recomender system  technique | RMSE | MAE | Time taken for predictions |
| collaborative | 0.20753485746047193 | 0.43209857080505715 | >1 hour |
| Collaborative along with base line approach | 0.6396424339363631 | 0.9045910050618966 | >1 hour |
| SVD | 1.0471340870910586 | 0.8392171832621713 | 134.5074987411499 |
| SVD with 90% energy | 1.047729841956572 | 0.8403502878605646 | 52.13040471076965 |
| CUR | 1.0473381541605242 | 0.8395303019330335 | 15.516500767236 |
| CUR with 90% energy | 1.0473387421235056 | 0.839529762041875 | 13.6386003493412253 |
| Latent factor model | 1,245233256126712 | 1.54423345339443 | 240.3429533412253 |