**Read Me- Assignment-4**

The program consists of 2 mappers, 2 reducers, driver class and a prediction.java class.

Input is downloaded from the link given. There are 2 files in the downloaded folder- u.data and u.item. u.data is given as an input to the 1st mapper. The functions of each of the classes are described as below.

**Mapper1:**  This class reads the input and creates the key value pair. Output of the mapper is in the format:

Key: userid

Value: movieid, rating

**Reducer1:** The output of the 1st mapper is given as the input to Reducer1. This class combines the movieid, rating for each user.

The output of this reducer is in the following format:

Key: userid

Value: Movieid, rating (in sorted order)

**Mapper2:** The output of the reducer-1 is given as the input to Mapper2. This class constructs 2 arrays- movie array, which contains the movieids and rating array which contains all the ratings. Then a loop is written to construct pairs of movies and pairs of ratings.

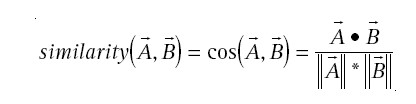
The output of this mapper is in the following format:

Key: movieid, movieid

Value: rating, rating

**Reducer2:** The output of the Mapper-2 is given as the input to Reducer-2. This class calculates the similarity between each movie by using cosine similarity.

The formula to calculate cosine similarity is-



The output of reducer-2 is in the following format:

Key: movieid, movieid

Value: similarity

**Driver Class**

The driver class sets the configuration for all the Mappers and Reducers through the Job configuration object. Hadoop uses those configurations to run the map-reduce programs. All file input and output paths are also defined in Driver class.

**Prediction. Java Class**

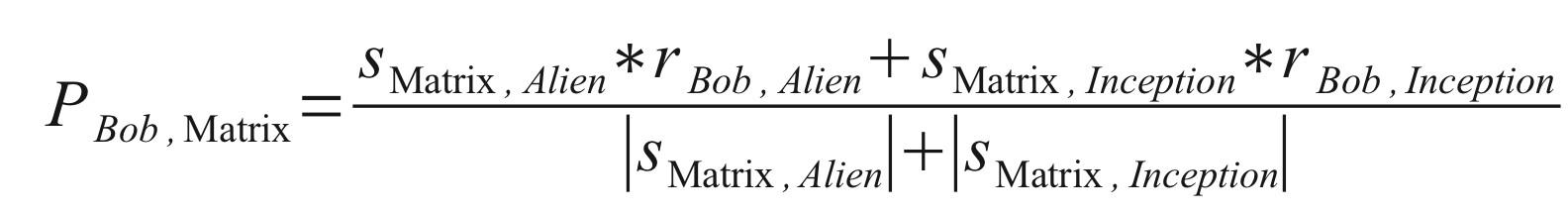
This class is used to calculate the predictions for a given userid. The main method takes 2 user inputs- userid and no. of movies to predict.

The following steps are followed to implement the feature.

1. Read the input data file to create a 2D matrix that stores the values in the format- moviematrix[userid][movieid] = rating.

2. Read the output of Reducer-2 and store the key-value pair in a hashmap. The hashmap should contain “movieid,movieid” as key and “similarity” as value.

3. Calculate the predicted rating for each user using the formula-

**

where predicted rating for matrix and alien is calculated for the user Bob.

4. Sort the array containing the predicted ratings. Run a loop to print top n recommended movies for the user.

Sample output:

Enter userid:

105

Enter no. of top movies:

5

Your top 5 recommended movies are:

1616

1556

1308

1307

1306

**Notes:**

1. Import the project to eclipse to run and see the output.

2. Change the file paths given in the program to your local directory before running the program.

3. All required Hadoop jar files must be in your library to execute the program.