INFORMATION RETRIEVAL (CS F469) ASSIGNMENT 2

Recommender System

Members

Rishi Bajargan 2015A4PS0422H

Krati Agarwal 2015B2A70778H

Mansi Ranka 2015A4PS0424H

Shreya Gupta 2015B2A70763H

**INTRODUCTION**

In this assignment, we have implemented three techniques of recommender system – Collaborative filtering, SVD and CUR decomposition, and compared their rms error, precision at top K, spearsman correlation and prediction time.

**ABOUT DATASET:**

Dataset is taken from (http://www.ieor.berkeley.edu/~goldberg/jester-data/) - dataset contains users and their ratings.

**DESIGN ARCHITECTURE:**

Below is the list of the algorithms used to implement recommender systems which have been implemented which include Collaborative filtering, Singular Value Decomposition and CUR.

Functionality Implemented

1. Collaborative Filtering (calculating similarity by users and predicting missing ratings)

2. Collaborative Filtering using global baseline approach.

3. SVD

4. SVD with 90% retained energy

5. CUR with sampling of rows and columns with replacement

6. CUR with sampling of rows and columns without replacement

**SOFTWARES/FRAMEWORKS USED:**

[python-Numpy,Pandas](https://pandas.pydata.org/pandas-docs/stable/install.html)

1. Install pip: sudo apt-get install python-pip

2. Install Numpy : sudo pip install -U numpy

3. Install pandas : sudo pip install pandas

Check for installation by Opening up a Python prompt by running the following:

python

At the prompt, type the following:

>>> import pandas

>>> import numpy

>>> print numpy.\_\_version\_\_

[python-xlrd module](https://www.loomio.org/) -

1. Install xlrd: pip install xlrd

2. Install xlwt: pip install xlwt

**PLATFORMS:**

To run recommender system make sure you got python installed.

To Run:

\* navigate to IR\_AS3 directory.

\* change path to the dataset in files to run

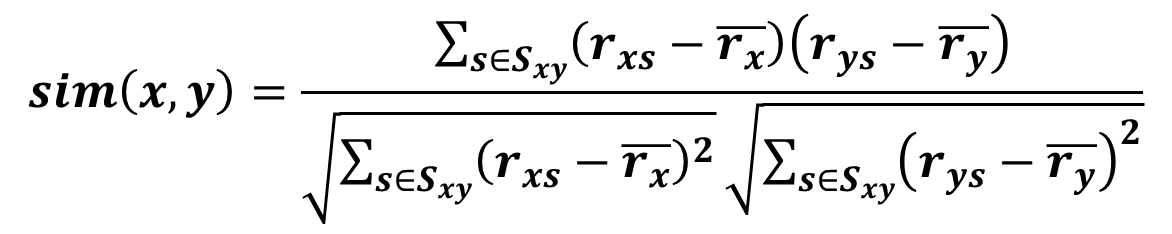
\* use python3 filename.

**FORMULAS, ASSUMPTIONS AND CONCEPTS USED**

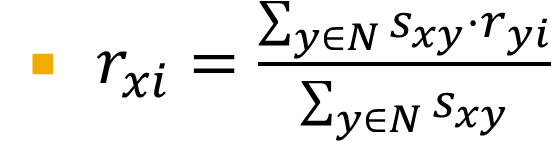
* For collaborative filtering, we have taken user-user similarity.

Formulas used

1. For calculating similarity of two users



1. For calculating rating of item i by user x



N= set of users whose similarity with x is greater than zero.

For collaborative filtering with baseline approach

Formulas used



* For SVD

For 90% retained energy, we have taken the largest singular values whose sum of squares are 90% of the total sum of squares.

* For CUR

We have taken randomly 25% of total number of users (approx. 420) rows and columns first including duplicates and then without duplicates.

**RESULT**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| TECHNIQUE | RMSE | PRECISION  AT TOP 50 | SPEARMAN  CORREALTION | TIME TAKEN (in sec) |
| Collaborative | 0.08 | 1.389 | 0.999 | 300 |
| Collaborative with baseline | 0.1581 | 1.440 | 0.999 | 180 |
| SVD | 0.60 | 1.53 | 0.9999 | 600 |
| SVD with 90% | 0.58 | 1.55 | 0.999 | 450 |
| CUR (with repetition) | 495.81 | 14.82 | 0.998 | 300 |
| CUR (without repetition) | 2260 | 568 | 0978 | 200 |