

Design Document
on
SVD and CUR Decomposition

Submitted on: -
21th November, 2016

Submitted to: -Dr. Aruna Malapati

Submitted by: -
Soamya Agrawal 2014A7PS185H
Ayushi Behl 2014A7PS145H
Tejeshwar Reddy 2014A8PS492H
Ruthvi Reddy 2014A7PS040H

[illegible]

The screenshot shows a Java IDE with the following components:

- Top Bar:** Displays the project name "IRRecommenderSystem - NetBeans IDE 8.1" and the current time "11:41 AM".
- Menu Bar:** Includes File, Edit, View, Navigate, Source, Refactor, Run, Debug, Team, Tools, Window, and Help.
- Toolbar:** Contains icons for file operations, navigation, and execution.
- Output Window:** Shows the execution results of the "IRRecommenderSystem (run)".
 - Matrix Output:** A large table of numerical values, likely representing a recommendation matrix. The first row contains values like 0.0180585486994278, 0.008353448678804765, etc.
 - Calculation Summary:** A section titled "Calculated finally the A matrix" followed by a large block of numerical data.
 - Success Message:** A message stating "BUILD SUCCESSFUL (total time: 6 seconds)".
- Bottom Bar:** Shows the "Output" tab and the current line number "28/25".

[illegible]

Error is 125.5647243456537

```
BUILD SUCCESSFUL (total time: 0 seconds)
```


About our Assignment:

- **Corpus:** <http://grouplens.org/datasets/movielens/> which includes user movie ratings and the movies are grouped by genre.
- **Data Structures:** We used HashMap for mapping Movie ID with the column's index of A matrix where A represents user-movie matrix. and Arrays for storing and for computational purposes
- Formulae used of SVD and CUR :
 - SVD->

SVD - Definition



$$A_{[m \times n]} = U_{[m \times r]} \Sigma_{[r \times r]} (V_{[n \times r]})^T$$

A: Input data matrix

– $m \times n$ matrix (e.g., m users, n movies)

U: Left singular vectors

– $m \times r$ matrix (m users, r concepts)

Σ : Singular values

– $r \times r$ diagonal matrix (strength of each 'concept')
(r : rank of the matrix **A**)

V: Right singular vectors

– $n \times r$ matrix (n movies, r concepts)

- CUR ->

CUR: How it Works



■ Sampling columns (similarly for rows):

Input: matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, sample size c

Output: $\mathbf{C}_d \in \mathbb{R}^{m \times c}$

1. for $x = 1 : n$ [column distribution]
2. $P(x) = \sum_i \mathbf{A}(i, x)^2 / \sum_{i,j} \mathbf{A}(i, j)^2$
3. for $i = 1 : c$ [sample columns]
4. Pick $j \in 1 : n$ based on distribution $P(j)$
5. Compute $\mathbf{C}_d(:, i) = \mathbf{A}(:, j) / \sqrt{cP(j)}$

Total length of all the columns

Note this is a randomized algorithm, same column can be sampled more than once

● Libraries :

- **Jama package**- We used Jama Package for all the matrix computations like finding rank, eigen values, eigen vectors etc.

Error Analysis + Time Analysis

For **SVD** -> The number of columns of A matrix are fixed as the number of distinct movies in the corpus and number of rows to be changing based on the users.

Using functions ,time and error were calculated.

For **CUR** -> The number of rows and number of columns for C and R matrix have been kept same. And for calculating error and time taken different values were passed.

SVD and CUR Error and Time Analysis ->

```
Java - IRRecommenderSystem/src/irrecommen/CompareSVDandCUR.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help

<terminated> CompareSVDandCUR [Java Application] C:\Program Files\Java\jre1.8.0_51\bin\javaw.exe (Nov 22, 2016, 12:16:45 AM)

SVD time is 15688
SVDError 16rows and 1000 columns 2.4760410374786667E-13

CUR time is 4
CURError 16rows and 16columns 4.2414797791426505E28
SVD time is 14810
SVDError 17rows and 1000 columns 2.6497167520727017E-13

CUR time is 3
CURError 17rows and 17columns 1.8038014947464489E52
SVD time is 14045
SVDError 18rows and 1000 columns 1.9241225506190257E-13

CUR time is 5
CURError 18rows and 18columns 594928.3011826843
SVD time is 13813
SVDError 19rows and 1000 columns 2.1223981209940558E-13

CUR time is 4
CURError 19rows and 19columns 2.7619303307857117E10
SVD time is 15283
SVDError 20rows and 1000 columns 3.7467456266808474E-13

CUR time is 3
CURError 20rows and 20columns 7.78001116756498E32
SVD time is 14475
SVDError 21rows and 1000 columns 3.1538113784490283E-13

CUR time is 4
CURError 21rows and 21columns 1.2537155758394135E55
```

```
Java - IRRecommenderSystem/src/irrecommen/CompareSVDandCUR.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help

<terminated> CompareSVDandCUR [Java Application] C:\Program Files\Java\jre1.8.0_51\bin\javaw.exe (Nov 22, 2016, 12:16:45 AM)

SVD time is 2025
SVDError 2rows and 1000 columns 1.6922795532425446E-15

CUR time is 54
CURError 2rows and 2columns 6.659973041975457
SVD time is 2754
SVDError 3rows and 1000 columns 3.00395245355628E-14

CUR time is 59
CURError 3rows and 3columns 32.065558096135064
SVD time is 17529
SVDError 4rows and 1000 columns 2.894792370107681E-14

CUR time is 36
CURError 4rows and 4columns 35.844339776012916
SVD time is 17722
SVDError 5rows and 1000 columns 1.3193246273719061E-13

CUR time is 5
CURError 5rows and 5columns 53.070152270920566
SVD time is 18473
SVDError 6rows and 1000 columns 6.370116010244474E-14

CUR time is 2
CURError 6rows and 6columns 56.69061826200273
SVD time is 16352
SVDError 7rows and 1000 columns 1.0650294511010298E-13

CUR time is 3
CURError 7rows and 7columns 57.616181758093155
SVD time is 17652
```