Design Document – IR Assignment 3

Team:

Sanjiv Yelthimar Shenoy - 2017A7PS0224H

Tulika Jha – 2017AAPS0364H

Manish Kumar Bachhu – 2017A7PS0036H

Sundeep Kumar Ammisetti – 2017A7PS1218H

Preprocessing

1. We created a rating matrix with movies as rows and users as columns.

Collaborative

1. We used the item-item collaborative model. For a given user in order to predict what the user will rate for a given item, we found all items similar to the given item. Then we took a weighted average of the user’s ratings for all the similar items in proportion to their similarity.
2. Some of the known ratings were hidden and used as test set. We found the mae, rsme error values and time taken based on the test data set.

SVD

1. We decomposed the ratings matrix into 3 matrices.
2. These were then multiplied to reconstruct the original matrix.
3. The error was then calculated and tabulated.

90% SVD

1. We took the required number of attributes such that the sum of their squares was greater than 90% of sum of squares of all diagonal elements in the sigma matrix.
2. The 3 matrices were then reduced accordingly.
3. They were then multiplied to get a matrix closer to the original matrix.
4. The error was then calculated and tabulated.

CUR

1. We decomposed the given matrix into 3 matrices.
2. The first had random columns. The columns were chosen according to higher probability by the frobenious norm.
3. The third matrix had random rows again based on the frobenious norm.
4. The second matrix was calculated by taking intersection of the random rows and columns and performing certain mathematical operations.
5. Finally, the matrices were multiplied and compared with the original.
6. The error was computed and tabulated.

CUR 90%

1. The Union matrix was reduced much that just greater than 90% covariance was maintained.
2. The matrices were then reduced accordingly.
3. The matrices were then multiplied to get a matrix closer to the original.
4. The error was then calculated and tabulated.

Latent Factor Model

1. The original matrix with rating even for the unknown values get got by using matrices Q and P
2. The matrices Q and P required were initialized with the values for svd decomposition.
3. Gradient Descent was then applied to make the model learn the generalization of the training set ratings.
4. The error was then calculated on the test set based on the difference between the actual and predicted values.
5. The error was then tabulated.

|  |  |  |  |
| --- | --- | --- | --- |
| **Recommender System Technique** | **Root Mean Squared Error (RMSE)** | **Mean Average Error (MAE)** | **Time taken for prediction** |
| Collaborative |  |  |  |
| Collaborative along with Baseline approach |  |  |  |
| SVD |  |  |  |
| SVD with 90 per cent energy |  |  |  |
| CUR |  |  |  |
| CUR with 90 per cent energy |  |  |  |
| Latent factor model |  |  |  |