Plagiarism Scan Report

0% Plagiarism	0% Exact Match
	0% Partial Match

100% Unique

Words	408
Characters	2778
Sentences	39
Paragraphs	38
Read Time	3 minute(s)
Speak Time	3 minute(s)

Content Checked For Plagiarism

1. Introduction
Recommender systems have become vital components of most contemporary applications, assisting users in dealing with huge information overload. This report analyses three methods of estimating the missing ratings in a recommendation system dataset:
* PCA with Mean-Filling
* PCA with Maximum Likelihood Estimation
* Eigenvalue and Eigenvector Analysis
The aim is to forecast missing ratings with precision and measure how good each approach is. These results showcase significance of matrix factorization approaches in sparcity of dataset.
2. Dataset Preparation
The dataset for this assignment was generated using the TMDB API in Assignment 1. Adjustments were made to represent real-world sparsity. Key steps included:
* Scaling Ratings: Adjusted all ratings to a 1-to-5 scale.
* Sparsity Analysis: Evaluated the sparsity of the dataset.
* Bias Detection: Identified potential biases in the dataset.
* Target Items: Selected two lowest-rated items as target items for prediction.
* Dataset Summary:
* Total Users (Tnu): 100
* Total Items (Tai): 20
* Number of Ratings: 1,200
The sparsity and bias analyses revealed a typical user-item matrix with approximately 60% missing ratings, reflecting challenges in real-world scenarios
3. Part 1: PCA Method with Mean-Filling
3.1 Methodology
1. Average Ratings: Calculated the mean rating for each target item.
2. Mean-Filling: Replaced missing values with corresponding mean ratings.
3. Covariance Matrix: Computed the covariance matrix.
4. Peer Identification: Determined top 5 and 10 peers for target items.
5. Prediction: Predicted missing ratings using reduced dimensions.
3.2 Results
* Predicted missing ratings for target items using top 5 and 10 peers.
* Mean-filling ensured computability but introduced bias.
Pros and Cons
* Pros: Simple and efficient.
* Cons: High bias in predictions, less accurate for sparse datasets.

4. Part 2: PCA Method with Maximum Likelihood Estimation

4.1 Methodology

1. Specified Entries:

Computed the covariance matrix using only the entries corresponding to specified ratings in the dataset. By focusing solely on the available data, the method avoided the distortion caused by imputing missing values arbitrarily.

2. Peer Identification:

Identified the top 5 and 10 peers for the target items based on their similarity, as derived from the covariance matrix. These peers were selected to provide meaningful context and more reliable predictions for the target item's missing ratings.

3. Prediction:

Predicted the missing ratings using a dimensionality reduction approach based on Principal Component Analysis (PCA). This method effectively captured the underlying structure of the data, ensuring that predictions were grounded in reduced but essential dimensions of variability.

Matched Source

No plagiarism found

Check By: Dupli Checker