Dupli Checker Date: 31-12-2024

Plagiarism Scan Report

2% Plagiarism	
0% Exact Match	
2% Partial Match	
98% Unique	
Words	913
Characters	6295
Sentences	59
Paragraphs	33
Read Time	5 minute(s)
Speak Time	7 minute(s)

Content Checked For Plagiarism

AIE425 Intelligent Recommender Systems, Fall Semester 24/25

Assignment #3: Dimensionality Reduction Methods

221100674, Nabila Ahmed Ali

1. Introduction:

Recommender systems play a vital role in tailoring user experiences by offering personalized suggestions based on preferences, behavior, and interactions. This report focuses on exploring advanced techniques in recommender systems through dimensionality reduction methods, aligning with the objectives outlined in the assignment. The goal is to analyze and implement key methods to improve the performance and accuracy of recommendations.

The assignment comprises three main parts, each centered on a specific approach to dimensionality reduction:

* Part 1: PCA Method with Mean-Filling

This section emphasizes the use of Principal Component Analysis (PCA) combined with mean-filling to

handle missing data. It involves generating covariance matrices, identifying top peers, and predicting ratings for selected target items based on reduced-dimensional representations.
* Part 2: PCA Method with Maximum Likelihood Estimation
Building on the foundation of PCA, this part introduces Maximum Likelihood Estimation (MLE) to compute the covariance between pairs of items. The focus is on predicting missing ratings while ensuring the effective transformation of data into reduced dimensions for more efficient computations.
* Part 3: Singular Value Decomposition (SVD) Method
The final section delves into Singular Value Decomposition (SVD), a powerful matrix factorization technique. By decomposing the ratings matrix, the task involves identifying latent factors that capture the essential structure of the data, reconstructing matrices, and predicting missing values with improved accuracy.
Throughout the report, the methods are analyzed in depth, with comparisons of their effectiveness and computational efficiency. By employing these techniques, this assignment highlights the practical applications of dimensionality reduction in recommender systems, paving the way for more robust and user-centric solutions.

1.1. Dataset Overview and Preparation:
The dataset consists of 884 rows and 4 columns: user_id, media_id, rating, and title. It represents user interactions with media content, where each record indicates a user's rating for a specific movie. The dataset contains:
* 60 unique users (user_id)
* 86 unique media items (media_id and title

1.2. Adjusting the Rating Scale:
The original ratings ranged from 1 to 10. For collaborative filtering, it is beneficial to normalize ratings to a smaller, standardized range, such as 1 to 5. This normalization improves consistency and makes it easier to compare user and item preferences.To achieve this, a linear scaling transformation was applied using the formula:Where:

* The target range (Scale Min and Scale Max) is 1 to 5.
* The original range (Min Original and Max Original) is 1 to 10.
The resulting ratings were rounded to two decimal places for accuracy. This adjustment ensured uniformity, eliminating any discrepancies caused by varying user rating scales. The transformed dataset was then saved as a CSV file to retain compatibility with further processing and analysis.

1.3. Total Users and Items:
The number of unique users (Tnu) and items (Tni) in the dataset was calculated to gauge its size and complexity. The outputs showed 60 unique users and 86 unique items, indicating a moderately sparse dataset suitable for collaborative filtering and dimensionality reduction techniques. The availability of numerous user-item interactions enables the calculation of reliable similarities, which form the foundation of effective CF systems.

1.4. Ratings per Product:
The step of counting ratings per product is an essential part of dataset preparation, offering critical insights into user interaction with various items in the dataset. Using Python, the number of ratings for each item, identified by its media_id, was calculated through grouping and aggregation techniques. Specifically, the dataset was grouped by media_id, and the total number of ratings for each unique item was determined. The results revealed significant variability, with some items, such as highly popular movies, receiving a large number of ratings, while others had minimal or no feedback at all. This variability underscored the sparsity of the dataset and the importance of applying dimensionality reduction techniques to improve the performance of recommendation algorithms.

1.5. Distribution of Ratings:

The histogram provides a visual representation of the distribution of ratings in the dataset, offering valuable insights into user preferences and potential biases. The x-axis represents the rating values, while the y-axis displays the frequency of each rating. Observing the histogram, it is clear that the majority of the ratings are concentrated between 8.0 and 9.0, with notable peaks around these values. This pattern suggests a positive bias in the dataset, where users predominantly gave high ratings to the items.
The relatively low frequency of ratings below 7.5 indicates that users are less inclined to assign lower ratings, potentially due to personal biases or the inherent quality of the items being rated. The near-symmetry around the central peak hints at a fairly consistent rating behavior among users, though the high clustering in the upper range could pose challenges for algorithms, as there is less variability to differentiate items effectively.
This distribution emphasizes the importance of preprocessing steps such as normalization and sparsity handling before applying dimensionality reduction techniques. By addressing these patterns, the analysis ensures a more balanced and unbiased approach in subsequent tasks. The histogram acts as a foundational step in understanding the dataset's characteristics, guiding the application of appropriate methodologies to improve recommendation accuracy.

1.6. Identifying the Two Lowest-Rated Items:

The two items with the lowest average ratings were identified and labeled as I1 and I2. These items were item 8587 and item 42269, representing products that might not appeal to a wide audience. Their inclusion as target items ensured a balanced evaluation of the recommendation algorithms.

1.7. Saving Results for Later Use:

Matched Source

Similarity 4%

Title:Singular Value Decomposition (SVD) | Advanced Matrix ...

Aug 16, 2024 — Singular Value Decomposition (SVD) is a powerful matrix factorization technique that breaks down any matrix into three components.

https://library.fiveable.me/advanced-matrix-computations/unit-2/singular-decomposition-svd/study-guide/s92yPJ316nBQ73XP