

Anime Recommendation System Using Cosine Similarity

1. Introduction

A recommendation system is an intelligent system that predicts user preferences and suggests items accordingly. In this project, a content-based recommendation system is implemented using cosine similarity to recommend anime titles based on their characteristics such as genre and rating.

2. Objective

The primary objective of this assignment is to design and implement a recommendation system that suggests similar anime titles using cosine similarity. The system analyzes anime features and computes similarity scores to generate meaningful recommendations.

3. Dataset Description

The Anime dataset contains information about various anime titles. Each record includes a unique ID, anime name, broadcast type (TV, Movie, OVA, etc.), genre information, number of episodes, average user rating, and the number of community members who rated the anime.

4. Data Preprocessing

Data preprocessing is a crucial step to ensure data quality. Missing values were identified and removed to avoid inconsistencies. Genre values were cleaned and transformed into a structured format suitable for machine learning algorithms.

5. Feature Extraction

Genres, being categorical in nature, were converted into numerical form using multi-label binarization. Numerical features such as ratings were normalized using Min-Max scaling to bring all features onto a comparable scale. These features together represent each anime in a multidimensional feature space.

6. Cosine Similarity

Cosine similarity measures the cosine of the angle between two vectors in a multidimensional space. It is widely used in recommendation systems as it focuses on the orientation of vectors rather than their magnitude. Higher cosine similarity values indicate greater similarity between anime titles.

7. Recommendation System Design

The recommendation system computes a cosine similarity matrix for all anime items. Given a target anime, the system retrieves the most similar anime titles based on highest similarity scores. Thresholds and top-N values can be adjusted to control the number of recommendations.

8. Evaluation

The dataset was split into training and testing sets to validate system performance. Evaluation metrics such as precision, recall, and F1-score help assess the relevance and accuracy of recommendations. The results indicate that content-based similarity provides consistent and interpretable recommendations.

9. Interview Questions and Answers

1. What is a recommendation system?

A recommendation system is a software tool that suggests items to users based on their preferences, behavior, or item similarities.

2. What is collaborative filtering?

Collaborative filtering recommends items by analyzing user-item interaction patterns and finding similar users or items.

3. Difference between user-based and item-based collaborative filtering?

User-based filtering finds similar users and recommends items liked by them, whereas item-based filtering finds similar items and recommends items related to those a user already liked.

4. What is content-based filtering?

Content-based filtering recommends items by comparing item features and suggesting similar items based on user preferences.

5. Why is cosine similarity used in recommendation systems?

Cosine similarity is effective for measuring similarity in high-dimensional data and is insensitive to magnitude differences.

6. What are the limitations of content-based recommendation systems?

They may suffer from limited diversity and over-specialization, recommending only items similar to those already consumed.

7. What is feature normalization and why is it important?

Feature normalization scales numerical values to a common range, ensuring no feature dominates the similarity calculation.

8. How can recommendation systems be improved?

They can be improved by combining collaborative and content-based methods (hybrid systems), using richer features, and incorporating user feedback.