Technical Documentation on Netflix Movie Recommendation System (Clustering)

Table of Contents

Introduction

The Netflix movie recommendation system is designed to provide personalized movie suggestions to users based on their preferences and behaviour. Clustering techniques are employed to group movies into categories that are relevant to a user's interests. This technical document outlines the process and methodologies used for clustering movies in the Netflix recommendation system.

Objective

The primary goal is to cluster movies into groups based on their attributes (genre, actors, ratings, etc.) to recommend similar movies to users who have shown an interest in a particular movie or category. This system aims to improve user satisfaction and engagement by providing relevant movie suggestions.

Dataset

The dataset used for this analysis contains information about movies available on Netflix. It includes features such as:

- **Movie Attributes**: Title, genre, release year, director, cast, etc.

- User Ratings: User ratings and reviews.

- Other Metadata: Duration, language, country, etc.

Data Preprocessing

- Data Cleaning: Handling missing values, duplicates, and ensuring data consistency.

- Feature Engineering: Extracting useful features from text data (e.g., actors, genres) for clustering.

- Vectorization: Converting textual information (like genres, cast) into numerical form (using techniques like TF-IDF or word embeddings).

Clustering Model

Feature Representation

- Transforming movie features into a format suitable for clustering, using techniques like TF-IDF, word embeddings (Word2Vec, GloVe), or other vectorization methods.

Clustering Algorithms

- K-Means Clustering: Partitioning movies into k clusters based on their attributes.

- Hierarchical Clustering: Creating a hierarchy of clusters to represent relationships between movies.

- DBSCAN or Other Density-Based Clustering: Identifying clusters based on data density.

Evaluation

- Intrinsic Metrics: Using metrics like silhouette score to evaluate the quality of clusters.

- Visual Inspection: Exploring clusters visually through dimensionality reduction techniques like PCA or t-SNE.

Recommendation System

- User-Movie Association

Associating users with movie clusters based on their interactions, ratings, and preferences.

- Providing Recommendations

Suggesting movies from the same cluster or similar clusters to the movies the user has interacted with or liked.

Conclusion

This technical document outlines the methodology used for clustering movies in the Netflix recommendation system. The clusters generated will be utilized to recommend similar movies to users, enhancing their viewing experience and engagement on the platform.

The clustering model and recommendation system aim to improve user satisfaction by offering personalized and relevant movie suggestions, ultimately contributing to increased user retention and engagement.

.