Web-Based Movie Recommendation System

Proposal

By

Adithya C

Arul Kumaran P

**Motivation**

In the era of information overload, pushing to users information matching their preference has been a valuable application and a major challenge in the field of data science. In this project, we proposed to explore the movie dataset, a dataset consisting of users’ rating and tagging towards movies, and implement a demo movie recommendation system.

**Intended final products**

* An algorithm/server that updates datasets by web scraping movie details often to corresponding datasets.
* A movie recommendation engine with a one line tag of compiled reviews from available review platforms.
* A HTML web page or dashboard giving movie recommendation

**Anticipated data sources**

* Movies 5000 dataset.
* Movie information scraped from web.
* Movie information scraped from twitter and social platforms.
* Movie information scarped from movie review web pages.

**Modules:**

* Module I : Developing Web Application.
* Module II : Creating database.
* Module III : Creating self-updating algorithim with API.
* Module IV : Creating Recommendation Algorithm

**Planned analyses/ visualizations / coding challenges**

**Planned analyses**

* Find popular movies and high-rating movies.
* Investigate the relationship among ratings and various factors, including movie genre, release year, rating time.
* Evaluate the similarity among movies.
* Analyze the changes in the ratings of the same movie every year, or the changes in the ratings of each category of movies, and try to get a rule through the analysis.
* Construct model to predict the rating of a specific movie. Preference on specific genre.
* Find the average rating after every year a movie published. And find the trend of that.
* Find the the average rating of each categories in different year.

**Coding challenges:**

* Create an algorithm to maintain datasets with latest movie details by web scraping.
* Convert collected-in-seconds timestamp into years
* Split and group movie by different genre
* Scrape movie basic information (length, director, actor, date…) from dataset for calculating movie-movie similarity
* Calculate movie-movie similarity based on movie description and user-generated tags
* Build models to calculate user-similarity

**Architectural Diagram:**

\ Home

Recommend Movie

Search (genres, actor name, director name, year, keyword)

Login with Google to fetch watch history

Query in database with the searched keyword

Content based filtering with watch history data.

Optimal result