COMP7240 Group Project Report

Hybrid Recommender no Yelp Dataset

BY GROUP JOJO 2024



Team Members

|  |
| --- |
| SUEN SHUI YAN 23435690 |
| WONG WAI HONG 23460407 |
| ZHANG HANYANG 23432985 |

# System Description

This recommender system is a sophisticated hybrid model that leverages a combination of content-based, collaborative filtering via Singular Value Decomposition (SVD), and collaborative filtering via neural networks to provide personalized recommendations. This system is designed to suggest businesses (like restaurants) to users based on their preferences and interactions.

## Main Functions

### Hybrid Approach

The system integrates content-based recommendations with two collaborative filtering methods (SVD and neural networks), offering a comprehensive recommendation strategy that capitalizes on the strengths of each method.

### Dynamic User and Item Profiles

It creates detailed profiles for both users and items (businesses) using the data from multiple datasets. These profiles are then used to match users with businesses that closely align with their preferences.

### Real-time Recommendations

The system updates its recommendations in real-time based on new user ratings, ensuring that the recommendations remain relevant and personalized.

### Explainability

For each recommended item, the system can provide explanations based on the contribution of each recommendation method, enhancing transparency and trustworthiness.

## Hybrid Recommendation Algorithm

### Content-Based Recommender

This module focuses on recommending items similar to what the user has liked in the past, based on item features such as categories or attributes of businesses.

### SVD Recommender

Utilizes Singular Value Decomposition for collaborative filtering, identifying latent factors from user-item interaction data to predict a user's preference for an item.

### NN Recommender

Employs neural networks to model complex non-linear relationships in the data, capturing deep patterns of user-item interactions.

The hybrid system combines these approaches to offset the limitations of individual methods (such as cold start problems and scalability issues) and to provide a more accurate and diversified set of recommendations. The hybridization design allows for leveraging content similarity, latent factor models, and deep learning insights simultaneously, offering a robust solution to various recommendation challenges.

## Datasets Used

The Yelp Dataset is a rich, publicly available dataset provided by Yelp for academic and learning purposes. It contains detailed information about local businesses, user reviews, and user interactions across many cities worldwide. By leveraging these datasets from the Yelp Dataset, the recommender system can perform complex analyses and predictions to offer highly personalized and contextually relevant business recommendations.

#### Business Dataset (`business.pkl`)

Includes comprehensive information about businesses listed on Yelp, such as business names, locations (latitude and longitude), categories (e.g., restaurants, bars, salons), and other attributes (e.g., Wi-Fi availability, parking, accessibility). This dataset enables the recommender system to identify and suggest businesses based on the user's location and preferences.

#### User Dataset (`user.pkl`)

Contains user profiles, including the user's review count, average rating given, and Yelp joining date. This dataset helps in understanding user behavior and preferences over time, crucial for tailoring personalized recommendations.

#### Review Dataset (`review.pkl`)

Comprises detailed reviews and ratings that users have left for businesses. Each review includes the user ID, business ID, stars (rating), and text content of the review. This rich dataset not only allows for analyzing user preferences but also helps in sentiment analysis and understanding the context behind ratings.

#### Photo Dataset (`photo.pkl`)

Contains mappings of photo IDs to businesses, providing a visual aspect to the recommendations. Photos can include images of the business, the services or products offered, and user-generated content. Incorporating visual elements into recommendations can enhance user engagement and provide additional information to assist users in making informed decisions.

# User Interface

The user interface for the hybrid recommender system is powered by Streamlit, an open-source app framework that is particularly well-suited for machine learning and data science projects. This interface enhances the system's accessibility and interactivity, allowing users to effortlessly tailor their recommendation experience based on their preferences and to visualize the recommendation outcomes

## Customizable Experience

Users can personalize their recommendation experience by selecting their profile, interests (categories), and preferred recommendation algorithms (Content-Based, Collaborative SVD, Collaborative NN) through a sidebar, offering a high degree of personalization.

A screenshot of a food app

Description automatically generated

## Interactive Recommendations

The main interface displays the top 10 recommended businesses, enriched with images, detailed ratings, and an option for users to rate these businesses, further tailoring the recommendations.

A screenshot of a computer

Description automatically generated

## Visual Insights

Incorporates both matplotlib and Plotly for generating interactive visualizations, such as a dining compass and feature match strength, providing users with deeper insights into why certain recommendations were made.

A blue dot on a white background

Description automatically generated

## Real-time Feedback Loop

Users can provide immediate ratings to the recommended businesses, which the system can use for real-time updates to recommendations, ensuring a dynamic and responsive user experience.

A screenshot of a website

Description automatically generated

## Explanatory Component

Offers explanations for each recommendation by visualizing the contributing factors from different recommendation algorithms, enhancing transparency and trust.

A screenshot of a graph

Description automatically generated

By leveraging Streamlit's capabilities, the hybrid recommender system not only delivers personalized and dynamic recommendations but also ensures an engaging and informative user experience. This interface serves as a bridge between the complex algorithms of the recommender system and the end-users, making advanced recommendation technologies accessible and understandable to a broad audience.

## References

### • Datasets:

* <https://www.yelp.com/dataset>

### • Development toolkits:

* <https://surprise.readthedocs.io/en/stable/index.html> (Python library

“Surprise” with collaborating filtering methods)

* <https://scikit-learn.org/stable/> (Python library "scikit-learn" for data

preprocessing, model evaluation, and other utilities)

* <https://streamlit.io/> (streamlit framework to build and share machine learning and data science web apps)
* <https://plotly.com/> (Python library for creating a wide variety of complex graphs)

### • Statistical methods:

* <https://en.wikipedia.org/wiki/Discounted_cumulative_gain>
* <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.ndcg_score.html>