Personalized Book Recommendation System

Professor: Bart P.G. Van Parys

Team BARtS: **B**ibhabasu Das, **A**bhranil Chakrabarti, **R**achi**t** Jain, **S**hreya Gupta

Description

The objective of this project is to build a "Book Recommendation System" that relies on user information on past books reviewed by them to give them recommendations on the next book to read, both at an overall and genre level. We make two-fold recommendations: **Titles to Read** and **Authors You May Like**.

Available Data

We rely on data retrieved from multiple data sources to get information on books, ISBN, author, review and corresponding reviewer information. The following data sources will be used:

- Amazon Data: 51.3M reviews and 2.9M products
    - **Sample Review**: ReviewerID, ProductID, Review, Overall Rating, Review Date, etc.
    - **Meta Data**: ProductID, Title, Price, Also_Bought, Also_Viewed, Bought_Together, Category, etc.
- Kaggle Data**:** 271k books, 279k Users, 1.1M Reviews
    - **Book Data Labels**: ISBN, Book-Title, Book-Author, Publisher, Year-of-Publication, Image-URL
    - **Rating Data Labels**: UserID, ISBN, Rating; **User Data**: User-ID, Location, Age
- Book Data with Genre: 207.5k books with Genres
- Goodreads Data: 10.3k books

Approach

We cater to our users by relying on their historical data available with us. We differentiate the following users:

- **New User:** There is no existing data for this user in terms of books read, bought or reviewed
- **Existing User:** We have some information about the user - books they have bought, read and some of their reviews

Stretch Goal: Implement a feature similar to "frequently bought together" when recommending a books.

For a new user, we try to gather some data by asking the user to rate some books, authors and their preferred genres, allowing the user to be treated like an existing user. When no data is available, we use a naive method.

- **Naive Methodology**: Recommend the most reviewed/positively reviewed in a given genre
- **Collaborative Filtering**: Collaborative filtering is based on the assumption that people who agreed in the past will agree in the future, and that they will like similar kinds of items as they liked in the past. The system generates recommendations using only information about rating profiles for different users or items
- **Content-Based Filtering**: Content-based recommenders treat recommendation as a user-specific classification problem and learn a classifier for the user's likes and dislikes based on an item's features. Content-based filtering methods are best suited to situations where there is known data on an item (name, location, description, etc.), but not on the user.

Evaluation: A comparative evaluation of the different techniques using metrics based on mean-squared error, precision or ranking measures (eg. nDCG) will be performed and estimating the improvement over a baseline (naive recommender)

Impact

The overall goal is to identify patterns in reading behavior across users to find similarities between them and rely on this to recommend a book to a user. This system is agnostic to product category and can be embedded as a part of customer experience improvement for any platform, thereby empowering the customer with reliable and personalized information.

Sources

- Case Study on Personalized Product Recommendation Engine
- Recommender system - Wikipedia | Product Recommendation System