BookMart: A Facebook Profile-Based Gift Book Recommender System

-Arushi Jain (2012130)
-Shweta Sood (2012164)

Problem Statement & Brief Overview

Recommendation systems help people make decisions in the complex information spaces. We focus on the question "What book should I gift to a friend on his/her birthday?" in the scope of this Project. BookMart is an attempt to provide recommendations for gifting books to your friends by tapping facebook data. We have tried to achieve a system that retrieves the important information about the user's friend without asking the user to remember or provide his friends' interests.

We analyze the Social Graph (friend's facebook public profile) and Interest Graph of books and apply machine learning algorithms to help person pick perfect book as a gift from hundreds of available options.

Methodology Used

For Book Recommendation we followed the following steps:

- **Data Collection** : We collected data from Facebook using its API with the help of an access token and user ID. The data included : UserID, User Name, Gender, List of Books read by user, Book ID
- **Data Extraction, Cleaning and Feature Extraction**:
  1. We generated excel files from the retrieved json data. We found the list of unique books from the books read by users. These books were categorized into many categories like - Detective, Comedy etc. This was done using stopwords. Eg- If a book contained stopwords like 'crime', 'thrilling', 'adventure', 'action', it was categorized into action category.
  2. Next, the list of all unique book names read by all users, had repetitions. We removed such repetitions and updated the entries of book names per user. Also, some book names retrieved were not actually books but were marked as book pages. These spurious books were removed. Also, authors for books were not retrieved from Facebook as this information was not available. Hence book authors were manually added.
  3. Finally we had the following 4 datasets : UserID, Book Names, List of Books with their Categories, List of Users with their gender and books read by them.
- **Algorithms** :
  1. Collaborative Algorithm
     a. Item-based Collaborative Filtering - If two items tend to have the same users like and dislike them, then they are similar and users are expected to have similar preferences for similar items. Item based similarity is deduced from user preferences rather than item metadata.
  2. Content Based Filtering - Algorithms that attempt to recommend items that are similar to items the user liked in the past. They treat the recommendation problem as a search for related items. Items selected for recommendation are items whose

content correlates the most with the user's preferences. Content based algorithms analyze item descriptions to identify items that are of particular interest.

3. Hybrid - For Hybrid algorithms, we thought of two main approaches overcoming shortcomings of both Collaborative and Content based Algorithms like- early rater problem, sparsity issue and opinions of users don't consistently match always.
   a. Collaborative Filtering and adding weightage to common categories of two books(Hybrid 1) - While calculating similarity of two books, we gave equal weightage to jaccard similarity score and number of common categories between two books.
   b. Content based filtering and collaborative filtering (Hybrid 2)- Gave equal weightage to similarity score of book and user and score calculated for a book to a user using collaborative filtering. Main idea behind hybridising the approaches were to overcome the shortcomings of both the approaches and trying to achieve better recommendation by combining the results achieved by both.

4. We did the same exercise by segregating the data based on male and female, such that a male user gets book recommended by observing the data of other male users and same for female.

Similarity measures Used

Cosine Similarity (Content based Filtering)

$$Similarity\ (User1, Book1) = \frac{dot\text{-}product\ (book\ categories\ preferred\ by\ user,\ category\ of\ book)}{||categories\ liked\ by\ user||\ .\ ||category\ of\ book||}$$

Jaccard Similarity (Item-Item based Collaborative Filtering)

$$Similarity\ (Book1, Book2) = \frac{Number\ of\ Common\ users\ between\ two\ books}{Union\ of\ Users\ for\ two\ books}$$

Results Obtained

In the case of Facebook profile data, ratings are unary: either a user likes an item (thereby making a positive association), or not. The absence of a 'like' can be interpreted either as a dislike or ignorance about the item. We therefore chose to use the Precision, because the unary rating scale makes the task of recommending books more like a classification problem. We define Precision as follows, where r_i denotes a book recommended to the user and L denotes the set of books liked by the user.

We have used Precision to calculate the final results

$$Precision = \frac{\sum_i \mathbf{1}\{r_i : r_i \in L\}}{\sum_i \mathbf{1}\{r_i\}}$$