DTI5125[EG] Data Science Applications Project Proposal:Book Recommendation System

Kareem Atif Mohamed Bakli[1], Gehad Hisham Hassanr[2], Kareem Khaled Moner Waly[3], and Mostafa Mahmoud Nofal[4]

[1]**kbakl031@uOttawa.ca**
[2]**gsoma101@uOttawa.ca**
[3]**kwaly008@uOttawa.ca**
[4]**mnofa091@uottawa.ca**

1 PROBLEM FORMULATION

A book recommendation system is a type of recommendation system where we have to recommend similar books to the reader based on his interest. The books recommendation system is used by online websites which provide ebooks like google play books, open library, good Read's, etc.Recommendation systems ought to increase profit from product sales. To achieve this, recommendations need to be relevant, novel and diverse.

2 METHODOLOGY :

1) **Content-Based Filtering**:

The algorithm recommends a book that is similar to the reader based on his interest. In simple words, Inthis algorithm, we try to find finding item look alike. For example, a person likes to watch history movies, so he may like reading history books too because the two items have similar tags. Only it looks similar between the content and does not focus more on the person who is watching this. Only it recommends the product which has the highest score based on past preferences. In the current approach, the cosine distance metric was the one used for item descriptions. Mathematically, it measures the cosine of the angle between two vectors projected in a multi-dimensional space.The process used to get the most similar items is detailed in Algorithm.

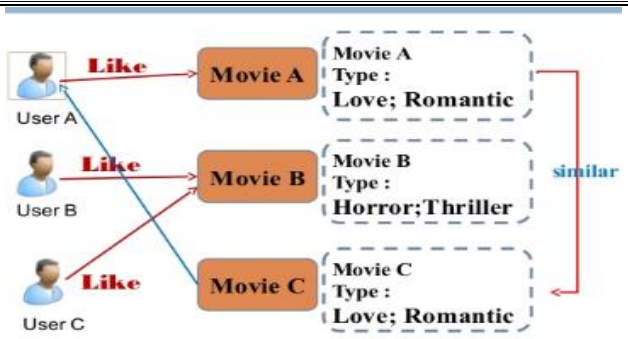

Fig1.Example:
Explain Content Based Filtering

2) Collaborative-based Filtering:

Collaborative based filtering recommender systems are based on past interactions of users and target items. In simple words here, we try to search for the look-alike customers and offer books based on what his or her lookalike has chosen. Let us understand with an example. X and Y are two similar users and Xuser has read A, B, and C books. And Y user has read B, C, and D books then we will recommend A book to Y user and D book to X user.

Fig1.Example:
Explain Collaborative Filtering

3) Hybrid Filtering Method:

It is basically a combination of both the above methods. It is a too complex model which recommendsproduct based on your history as well based on similar users like you.
There are some organizations that use this method like Facebook which shows news which is importantfor you and for others also in your network and the same is used by Linkedin too.

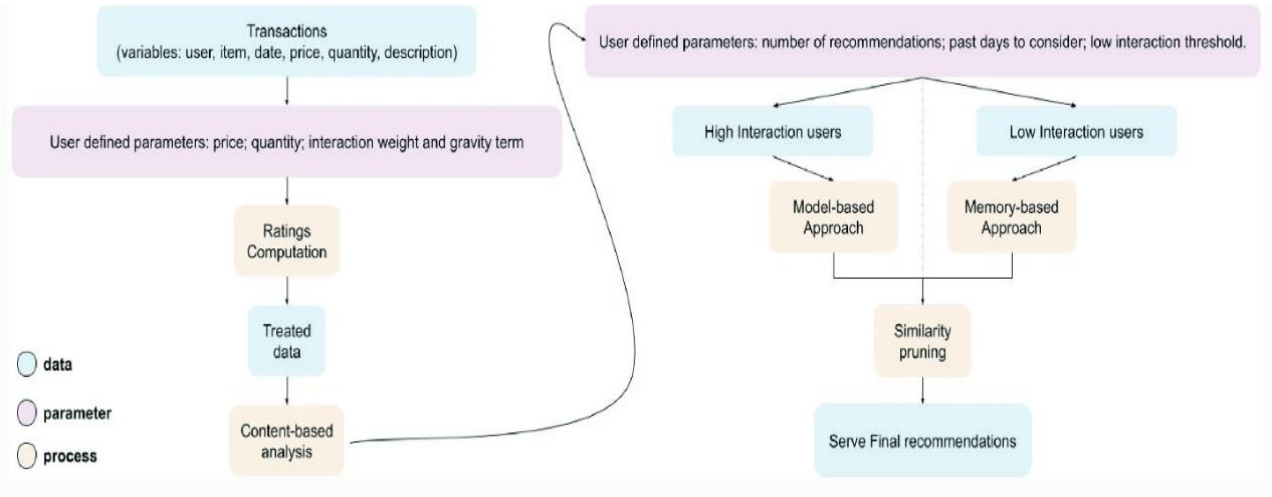

Fig3 Example of hybrid Method

DATA DESCRIPTION AND DATA SOURCES:

Book-Crossing Dataset:

Collected by Cai-Nicolas Ziegler in a 4-week crawl (August / September 2004) from the Book-Crossing community with kind permission from Ron Hornbaker, CTO of Humankind Systems. Contains 278,858users (anonymized but with demographic information) providing 1,149,780 ratings (explicit / implicit) about 271,379 books.

-Dataset Description:

we have 3 csv files in our dataset which is extracted from some books selling websites.1-Books:
first are about books which contain all the information related to books like an author, title, publicationyear, etc.
2-Users:
The second file contains registered user's information like user id, location.3-ratings :
Ratings contain information like which user has given how much rating to which book.
So based on all these three files we can build a powerful collaborative filtering model. let's get started.

3 EVALUATION METHOD/S:

1-Mean Average Precision:

Using MAP to evaluate a recommender algorithm implies that we are treating the recommendation like a ranking task. This often makes perfect sense since a user has a finite amount of time and attention and we want to show the top recommendations first and maybe market them more aggressively. In recommendation systems MAP computes the mean of the Average Precision (AP) over all your users.

The AP is a measure that takes in a ranked list of your N recommendations and compares it to a list of the true set of ”correct” or ”relevant” recommendations for that user. AP rewards you for having a lot of ”correct” (relevant) recommendations in your list, and rewards you for putting the most likely correct recommendations at the top (you are penalized more when incorrect guesses are higher up in the ranking). So order of ”hits” and ”misses” matters a lot in computing an AP score, but once you have front-loaded your best guesses you can never decrease your AP by tacking on more

.

2-Coverage:

Coverage is the percent of items in the training data the model is able to recommend on a test set

.

3-Personalization:

Personalization is a great way to assess if a model recommends many of the same items to different users. It is the dissimilarity (1- cosine similarity) between user’s lists of recommendations. An example will best illustrate how personalization is calculated

```
example_predictions = [
    ['A', 'B', 'C', 'D'],
    ['A', 'B', 'C', 'X'],
    ['A', 'B', 'C', 'Z']
]
```

Example list of recommended items for 3 different users.

	A	C	B	D	X	Z
0	1	1	1	1	0	0
1	1	1	1	0	1	0
2	1	1	1	0	0	1

First, the recommended items for each user are represented as binary indicator variables (1: the item was recommended to the user. 0: the item was not recommended to the user).

```
[1.   , 0.75, 0.75]
[0.75, 1.   , 0.75]
[0.75, 0.75, 1.   ]
```

Then, the cosine similarity matrix is calculated across all user’s recommendation vectors.

```
personalization = 1-0.75 = 0.25
```

Finally, the average of the upper triangle of the cosine matrix is calculated. The personalization is 1-the average cosine similarity.

4 RESULTS EXPECTATIONS:

Book Recommendation system has the ability to predict whether a particular user would prefer which book based on the user's interest. Recommender systems are beneficial to both service providers and users.They reduce transaction costs of finding and selecting books in an online shopping ebook environment.