Recommender systems rely on various data mining techniques for information

retrieval [9]. The one criterion that is used to classify them is the filtering

technique. There are five common techniques—Collaborative Filtering, Content-

Based, Knowledge-based, Group recommender systems, and Hybrid [17]. (Direct Citation)

**Collaborative Filtering**

It is a technique used to generate product/service recommendation to a user based on the preferences of other users with similar attitude. So, let’s assume that you have decided to buy a car, then this method analyses your behavior and matches it with other users with similar interest, then recommend a car that those users like since you have same interest as others.

It analyses your behavior, then it matches it with other users with same interest, then recommend you a product that others liked due to similar taste. If you want to buy a car then the system will recommend a car based on other people who has the same taste as you and bought that car in the past. So, if you have a similar taste as your cousin then chances are that you buy another product that your cousin bought in the past, so if your cousin liked the product, then you will like it as well because you share same interest or are like minded.

The system uses statistical methods to search for a set of users who have similar transactions history to the active user. (2)

From Textbook:

Collaborative Filtering is a technique used to make automatic predictions about the preference of a user derived by analyzing the preferences of other users that have similar interests. The underlined characteristic of this type of filtering is that if a person A has the same preference as person B, therefore, person A will be more likely to like another service/product that person B likes comparing to a random customer.

It can further split into six main categories

* model-based,
* memory-based,
* clustering based,
* matrix and tensor factorization,
* deep learning, and
* hybrid method.

**Memory-Based/Neighborhood Collaborative Filtering**

This Model uses statistical relationship such as correlation related to user’s rating to make recommendations. This approach uses the entire dataset to successfully recommend items.

Diagram

Description automatically generated

Two types:

* User Based
* Item Based

**User Based:**

It is a recommendation technique that recommends item to a certain buyer based on the ratings of that item by other people with similar interest as the buyer. So, pretty much finds users that have similar taste as you and make a recommendation based on the highest rated item voted by those users.

**Item Based:**

A rating is assigned to an item that you wish to buy based on your ratings of similar products that you have already rated.

**Model Based Collaborative Filtering**

Comparing to the memory-based approach, the model-based one usually does not use

the entire dataset to compute the predictions, but it uses a training set for predicting the future ratings. For recommending an item, a model first analyzes the user-item matrix. Next, it identifies the relationship among the items, and then it compares these relations with top-N recommendations list. After a model is constructed, predictions could be computed very fast. (1) (direct citation)

Making recommendations with this approach involves a part of the entire dataset. So, a subset of the entire dataset is taken to build a model, which is used to generate recommendations.

Diagram

Description automatically generated

1. **Explicit:** These ratings are the specific rating that a user gives to a product (for example, a user rates a book 3 on a scale of 1 to 5). These explicit ratings are directly used in the extractions of users’ interest for future recommendation. The disadvantage of explicit data is that it makes user responsible for data collection and future rating prediction who hardly takes interest to give a rating on a particular item. (Direct Citation)
2. **Implicit:** These ratings are collected by logging the user’s data generated while browsing the website. Implicit data are easier to collect as it does not put any pressure on the user to rate the products on the site. However, dealing with an implicit rating is very complicated as it is hard to find the users’ preferences from these collected users’ browsing data. Using these collected ratings (explicit or implicit); RSs predict the unknown ratings of the user based on different similarity metrics and these predicted ratings used in the recommendation process.(Direct Citation)

Text

Description automatically generated

Graphical user interface, text

Description automatically generated

Text, letter

Description automatically generated

Text, letter

Description automatically generated

Text, letter

Description automatically generated

Text, letter

Description automatically generated

Very Important Datebases

Text

Description automatically generated

Iibraries:

$ conda install -c conda-forge scikit-surprise

References:

1. (Zbigniew W. Ras (editor), Alicja Wieczorkowska (editor), Shusaku Tsumoto (editor), Recommender Systems for medicine and Music 2021)
2. Comparative-analysis-memory-based-model-based-recommendation-systems