Book Recommendation System

Sourabh Pramanik
Data science trainee,
AlmaBetter, Bangalore

Abstract:

Machine learning is a scientific study of statistical model and algorithms. In this research I will use the machine learning algorithms, K-NN and matrix factorization. In the books recommendations system BX books dataset is used. Suggestion method is a selection strategy which was used for collective selection and material-based sorting strategies. Pattern filtering technique is carried to suggest a consumer to an element the "rank" or "first option." Suggestion process collected information was about either the customer's first option on unusual subject relevant to films, books, travel, TV and commerce, etc. And from the other side, an effective selection of books recommendation system design utilizes prior scores or background of the customer. Cooperative sorting is a process of measuring and processing the categories across user opinions. Cooperative filtering first gathers the rankings or a preference of books provided by multiple users and then suggests books to different individuals based on various previous tastes and preferences. K-Means Multipathing together with K-Nearest Neighbor is applied on the BX dataset to achieve the greatest-optimized outcome.

1. Problem Statement

In today's world there are lots of books present. Users read books and give ratings and reviews according to their interest and experience.

In this project we will create a recommendation system which will provide genuine suggestion of the books to the user.

Data Description

In this project we have 3 datasets-

1. Books
    - ISBN
    - BOOK TITLE
    - BOOK AUTHOR
    - YEAR OF PUBLICATION
    - PUBLISHER
    - IMAGE URL
2. Users
    - USER ID
    - LOCATION

- AGE

3. ratings
   - USER ID
   - ISBN
   - BOOK RATING

2. Introduction

Recommendation system is defined as the computer program that helps the user determine goods and content by predict the users rating of each item and presentation them the substance that they would rate highly. The recommendation system is containing three types that are: collaborative filtering, content based filtering and hybrid filtering.

- Collaborative filtering

The collaborative filtering is based on customer's behaviors' and past likes. It is used to make automatic predictions, filtering collecting preferences, interest of a user and taste information of many users. .Collaborative filtering can be performed in two ways mainly, model-based and memory-based.

- Memory based

In memory-based collaborative filtering, the whole dataset is used to make a recommendation. 1.3 Model based This method does not use the complete dataset for generating recommendations. The model then makes recommendations on the test data. There are three steps in collaborative filtering recommendation. 1) Established a user model. 2) Find the nearest neighbor user 3) finally (Gao, 2019)generates a recommendation list.

We constructed a book recommender using collaborative item. For this aim, we used an open source recommendation system package for Python. The other frameworks we used were Pandas; for handling large datasets, SciKit learns for running the K-nearest neighbors algorithm and NumPy for its data structure capabilities. Our choice of programming language was Python.

3. Objective of our Project-

Building a Book recommendation system which can suggest books matching with user's interest.

4. Steps involved:

- **Exploratory Data Analysis**
  After reading the data we will perform Exploratory Data analysis.
- **Null values Treatment**
  Only 'Rating' dataset contain some null values in 'Age' column.
- **Visualization**
  We have done some visualization of Rating counts and Age group of users.

- **Model**

  We will use the Nearest Neighbour algorithm which is the similar as K nearest which is used for clustering based on Euclidian distance.

	userID	ISBN	bookRating	bookTitle	totalRatingCount	Location
0	277427	002542730X	10	Politically Correct Bedtime Stories: Modern Ta...	82	gilbert, arizona, usa
1	3363	002542730X	0	Politically Correct Bedtime Stories: Modern Ta...	82	knoxville, tennessee, usa
3	12538	002542730X	10	Politically Correct Bedtime Stories: Modern Ta...	82	byron, minnesota, usa
4	13552	002542730X	0	Politically Correct Bedtime Stories: Modern Ta...	82	cordova, tennessee, usa
5	16795	002542730X	0	Politically Correct Bedtime Stories: Modern Ta...	82	mechanicsville, maryland, usa

  In this project we are using only the records of US and Canada people .Using only those books whose rating greater than 50.

5. Modelling-

Here we are using Collaborative Filtering Technique.

Collaborative Filtering-

In the newer, narrower sense, collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating).

We have prepared our dataset for modelling. we are using the Nearest Neighbours algorithm which is the similar as K nearest which is used for clustering based on Euclidian distance.

Nearest Neighbor Algorithm-

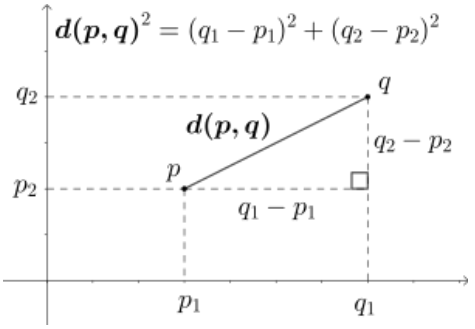
Nearest Neighbours is a simple algorithm widely used in predictive analysis to cluster data by assigning an item to a cluster by determining what other items are most similar to it. A typical use of the Nearest Neighbours algorithm follows these steps: Derive a similarity matrix from the items in the dataset.

Euclidian Distance –

The Euclidean distance between two points in Euclidean space is the length of a line segment between the two points.
It can be calculated from the Cartesian coordinates of the points using the Pythagorean theorem, therefore occasionally being called the Pythagorean distance. These names come from the ancient Greek mathematicians Euclid and Pythagoras, although Euclid did not represent distances as numbers, and the connection from the Pythagorean theorem to distance calculation was not made until the 18th century.

The distance between two objects that are not points is usually defined to be the smallest distance among pairs of points from the two objects. Formulas are known for computing distances between different types of objects, such as the distance from a point to a line. In advanced mathematics, the concept of distance has been generalized to abstract metric spaces, and other distances than Euclidean have been studied. In some applications in

statistics and optimization, the square of the Euclidean distance is used instead of the distance itself.



Here we are specifying an algorithm which is **brute** means find the distance of every point to every other point. And we will specify "**metric=cosine**" so that the algorithm will calculate the cosine similarity between rating vectors. Finally, we are fitting the model.

Cosine Similarity-

Cosine similarity is a metric used to measure how similar the documents are irrespective of their size. Mathematically, it measures the cosine of the angle between two vectors projected in a multi-dimensional space. The cosine similarity is advantageous because even if the two similar documents are far apart by the Euclidean distance (due to the size of the document), chances are they may still be oriented closer together .The smaller the angle, higher the cosine similarity.

Conclusion -

In this project we have made the Recommendation system which use Collaborative Filtering .The main goal was speed of recommendation .Experimental result shows that the proposed method shows relevant Recommendations.

The proposed methods can be applied in other domain also like Movies, music, Videos etc.

References-

1. MachineLearningMastery
2. GeeksforGeeks
3. Analytics Vidhya
4. TowardsDataScience
5. Stack Overflow