International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Book Recommendation Using K Nearest Neighbour and Collaborative Filtering

Snehitha S[1], Adithya R[2], Dr. Sujithra M[3]

[1]PG student, Department of Computing, Coimbatore Institute of TechnologyCoimbatore Institute of Technology, Coimbatore,India
[2]Assistant Professor,, Department of Computing, Coimbatore Institute of TechnologyCoimbatore Institute of Technology, Coimbatore,India

ABSTRACT:

With the increase in demand of items amongst customer enhances the growth in information technology and ecommerce websites. This demand is increased due to the availability of web services Personalized preferences and decision making are generated in an application called Recommendation system using an information filtering technique. Relevant features and related items are the characteristic on which this technique works. Suggestion of items, according to user preferences is most important, so suggestion according to similarities provides suitable recommendation. The working of recommendation system for administration has been researched in recent years.On the basis of price, quality, publisher and author, recommendation is done. Academician and students are more preferred by the book recommendation system. The purpose is to work on the profile of students and according to the interest based on the store profile recommend items at the time of searching, based on the search and interest of user. Variety of books are offered by book recommendation system, it display the results based on the search of user.

Keywords: RecommenderSystems, Collaborative filtering, K Nearest Neighbour

1.INTRODUCTION

The library is fundamental in an educational institution such as a university because of the library as a source of media for storing literature in conducting research and teaching activities. Each book has a relationship with other books in terms of content and references. A large number of books will contain a lot of data information as well,So, a search system is needed so that it is easy to search. Book search system based on the title, author,publisher, and book subject still uses the

syntax method which is based on the basic search. The use of the syntax method will contain a lot of information and is not effective in searching books that fit the user's needs.

Educational institutions that have a book search system in a library are still using a system with the syntax method that can cause a buildup of information so that users will long decide on the choiceof books to borrow. The problem of information accumulation will have an impact on the effectiveness of the time and quality of references because it can cause human errors in selecting books. A good system is a system that can simplify work and can work more effectively to support work acceleration without causing problems, it is necessary to design a book search recommendation system that can be a solution to problems that occur. A book search recommendation system is needed to reduce large data output so that book search will be more effective compared to book search system with syntax method. The book search recommendation system uses a user-based collaborative filtering method. the method uses the opinions of a user community to help an individual from the community to find certain information.

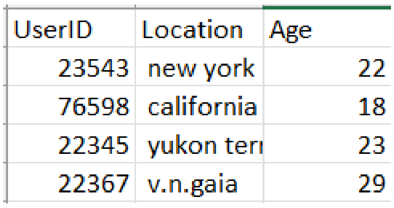
The formulation of the problem in this research is the development of library information systems using the User-Based Collaborative Filtering method as a recommendation for searching for books in the library. The Specific Purpose of this Research is to develop a book search recommendation system using UserBased Collaborative Filtering.

2.DATASET:

The Book-Crossing dataset comprises 3 files.

Users:

Contains the users. Note that user IDs (User-ID) have been anonymized and map to integers. Demographic data is provided (Location, Age) if available. Otherwise, these fields contain NULL-values.

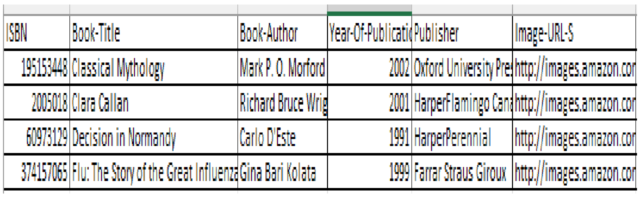


UserID	Location	Age
23543	new york	22
76598	california	18
22345	yukon ter	23
22367	v.n.gaia	29

Figure 2.1 Users

Books:

Books are identified by their respective ISBN. Invalid ISBNs have already been removed from the dataset. Moreover, some content-based information is given (Book-Title, Book-Author, Year-Of-Publication, Publisher), obtained from Amazon Web Services. Note that in case of several authors, only the first is provided. URLs linking to cover images are also given, appearing in three different flavours (Image-URL-S, Image-URL-M, Image-URL-L), i.e., small, medium, large. These URLs point to the Amazon web site.

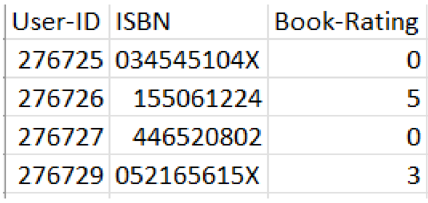


ISBN	Book-Title	Book-Author	Year-Of-Publicatio	Publisher	Image-URL-S
195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Pres	http://images.amazon.cor
2005018	Clara Callan	Richard Bruce Wrig	2001	HarperFlamingo Cana	http://images.amazon.cor
60973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	http://images.amazon.cor
374157065	Flu: The Story of the Great Influenza	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.amazon.cor

Figure 2.2 Books

Ratings:

Contains the book rating information. Ratings (Book-Rating) are either explicit, expressed on a scale from 1-10 (higher values denoting higher appreciation), or implicit, expressed by 0.



User-ID	ISBN	Book-Rating
276725	034545104X	0
276726	155061224	5
276727	446520802	0
276729	052165615X	3

Figure 2.3 Ratings

3.METHODOLOGY:

The method used to create a recommendation system is collaborative filtering and will use centered cosine similarity, cosine similarity, and k nearest neighbors. In giving a book recommendation to students in implementing the program code there are a number of steps that make all the data from borrowing the book into a matrix in the form of an array that will be calculated using an algorithm, in the second stage will be done the process of normalizing the ranking because in making book recommendations using borrowing data the number of books will continue to grow if borrowed by the same student and this makes the ranking scale erratic so that it needs normalization of ranks using centered cosine similarity, after the stage of normalization ranking is complete it will start using the cosine similarity algorithm to get the results of book rating, for the last stage sort the results of the previous calculation from the closest to k nearest neighbors (KNN).

3.1 Cosine Similarity

In this step, the results of previous calculations that have been obtained with Centered Cosine Similarity will then be replaced by the NaN value with the average value of the item.

After there is no NaN value, it will proceed with calculations using the Cosine Similarity algorithm. The library needed to calculate cosine similarity is sklearn.

Cosine similarity ranges from -1 to 1. 1 indicates the items are the same whereas -1 represents the compared items are dissimilar. Note that the cosine similarity in case (b) is 1 (similar) though the size of the blue jet ski vector is higher than the orange one. This indicates that cosine similarity is independent of the magnitude or the size of the vectors. It depends only on the direction of the vectors. When the angle between the vectors is small, the similarity is higher.

To get similar items to use the Cosine Similarity algorithm with the formula:

$$\text{Cosine similarity} = \frac{a.b}{||a||\ ||b||}$$

The results obtained from the implementation of using the Cosine Similarity algorithm using the librarysklearn.metrics.pairwise.cosine_similarity and displayed with library pandas in tabular form are as follows. where, a and b are the vectors of items a and b, and ||a|| and ||b|| are the Euclidean norm or the magnitude of the vector. This formula is more apt for machine learning applications, where we know the values of the vectors.

3.2 Collabaorative filtering:

This method matches the people of similar taste and then on the basis of personalized recommendation, recommend the user. This algorithm is classified into two entities, the user entity and the item entity. The user entity works on the basis of rating, they rate the item according to their opinion about that item. Recommender system mainly uses collaborative filtering or the combination of it with other algorithm. It mainly focuses on user with same preference and taste and suggest items to them on the basis of selection of items by those users.

Collaborative methods for recommender systems are methods that are based solely on the past interactions recorded between users and items in order to produce new recommendations. These interactions are stored in the so-called "user-item interactions matrix".

The class of collaborative filtering algorithms is divided into two sub-categories that are generally called memory based and model based approaches. Memory based approaches directly works with values of recorded interactions, assuming no model, and are essentially based on nearest neighbours search (for example, find the closest users from a user of interest and suggest the most popular items among these neighbours).

Model based approaches assume an underlying "generative" model that explains the user-item interactions and try to discover it in order to make new predictions.

The main advantage of collaborative approaches is that they require no information about users or items and, so, they can be used in many situations. Moreover, the more users interact with items the more new recommendations become accurate: for a fixed set of users and items, new interactions recorded over time bring new information and make the system more and more effective.

Collaborative filtering is used to tailor recommendations based on the behavior of persons with similar interests. Sometimes it can be based on an item bought by the user. Since this method does not require a person himself to always contribute to a data store, and voids can be filled by the actions of other persons/ actions by the same person on other items.

3.2.1 Memory-based:

User-User:

In order to make a new recommendation to a user, user-user method roughly tries to identify users with the most similar "interactions profile" (nearest neighbours) in order to suggest items that are the most popular among these neighbours (and that are "new" to our user).

This method is said to be "user-centred" as it represent users based on their interactions with items and evaluate distances between users.Assume that we want to make a recommendation for a given user. First, every user can be represented by its vector of interactions with the different items ("its line" in the interaction matrix). Then, we can compute some kind of "similarity" betweenour user of interest and every other users. That similarity measure is such that two users with similar interactions on the same items should be considered as being close. Once similarities to every users have been computed, we can keep the k-nearest-neighbours to our user and then suggest the most popular items among them (only looking at the items that our reference user has not interacted with yet).

Item-Item:

To make a new recommendation to a user, the idea of item-item method is to find items similar to the ones the user already "positively" interacted with. Two items are considered to be similar if most of the users that have interacted with both of them did it in a similar way. This method is said to be "item-centred" as it represent items based on interactions users had with them and evaluate distances between those items. Assume that we want to make a recommendation for a given user. First, we consider the item this user liked the most and represent it (as all the other items) by its vector of interaction with every users ("its column" in the interaction matrix). Then, we can compute similarities between the "best item" and all the other items. Once the similarities have been computed, we can then keep the k-nearest-neighbors to the selected "best item" that are new to our user of interest and recommend these items.

The collaborative nature of this approach is apparent when the model learns the embeddings. Suppose the embedding vectors for the movies are fixed. Then, the model can learn an embedding vector for the users to best explain their preferences. Consequently, embeddings of users with similar preferences will be close together. Similarly, if the embeddings for the users are fixed, then we canlearn movie embeddings to best explain the feedback matrix. As a result, embeddings of movies liked by similar users will be close in the embedding space.

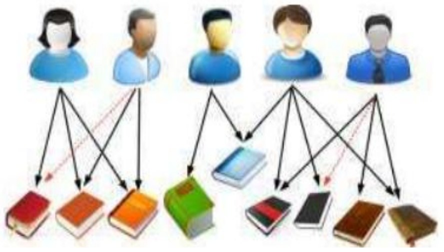


Figure3.2.1.1Collaborative Filtering for Book Recommendation

3.3 K Nearest Neighbors (KNN):

The standard method of Collaborative Filtering is known as Nearest Neighborhood algorithm. There are user-based CF and item-based CF. Let's first look at User-based CF.

Where it has an n × m matrix of ratings, with user u_i, i = 1, …n and item p□ , j=1,…m. Now we want to predict the rating r_i□ if target user i did not watch/rate an item j. The process is to calculate the similarities between target user i and all other users, select the top X similar users, and take the weighted average of ratings from these X users with similarities as weights.The next stage is to find other members who are closest to the members who want to find book recommendations and rank them. In this study, looking for 10 members who are most similar to members who want to find book recommendations. The results obtained and displayed in tabular form using the Pandas library are as follows.

KNN is a machine learning algorithm to find clusters of similar users based on common book ratings, and make predictions using the average rating of top-k nearest neighbors. For example, we first presentratings in a matrix, with the matrix having one row for each item (book) and one column for each user.

Then the table is converted to a 2D matrix, and fill the missing values with zeros (since distances between rating vectors will be calculated). Then transformed the values(ratings) of the matrix dataframe into a scipy sparse matrix for more efficient calculations.

Unsupervisedalgorithms were used with sklearn.neighbors. The algorithm, used to compute the nearest neighbors is “brute”, and we specify “metric=cosine” so that the algorithm will calculate the cosine similarity between rating vectors. Finally, we fit the model. Then the kNN algorithm measures distance to determine the “closeness” of instances. It then classifies an instance by finding its nearest neighbors, and picks the most popular class among the neighbors.

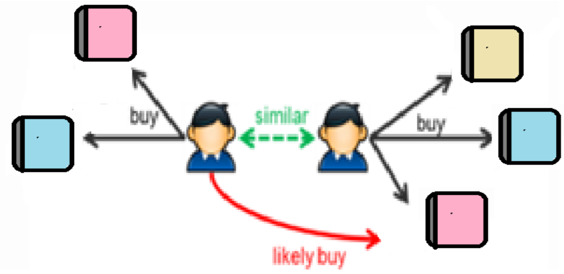
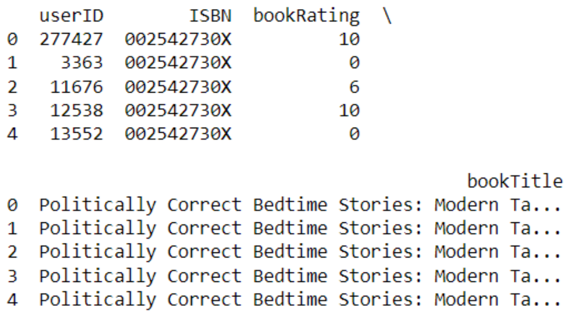


Figure 3.2.1KNN for Book Recommendation

4. EXPERIMENTAL RESULTS:

With the use of collaborative filtering and KNN, the top five results based on users and items are retrieved.

Here Figure 4.1 shows the book titles retrieved from the response of user’s reviews.

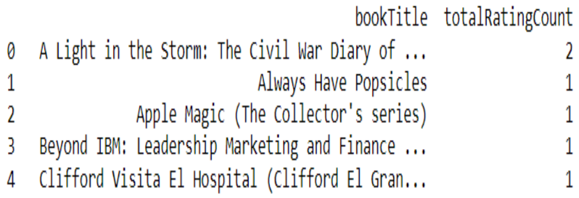


```
   userID       ISBN  bookRating  \
0  277427  002542730X         10
1    3363  002542730X          0
2   11676  002542730X          6
3   12538  002542730X         10
4   13552  002542730X          0

                                          bookTitle
0  Politically Correct Bedtime Stories: Modern Ta...
1  Politically Correct Bedtime Stories: Modern Ta...
2  Politically Correct Bedtime Stories: Modern Ta...
3  Politically Correct Bedtime Stories: Modern Ta...
4  Politically Correct Bedtime Stories: Modern Ta...
```

Figure 4.1 User’s corresponding book title

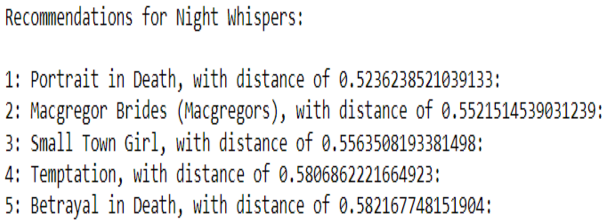
Figure 4.2 shows the grouping by book titles and create a new column for total rating count.



```
                                           bookTitle  totalRatingCount
0  A Light in the Storm: The Civil War Diary of ...                 2
1                                 Always Have Popsicles                 1
2                    Apple Magic (The Collector's series)                 1
3  Beyond IBM: Leadership Marketing and Finance ...                 1
4  Clifford Visita El Hospital (Clifford El Gran...                 1
```

Figure 4.2 Books with total ratings

In figure 4.3, The books related to the requirement that has the highest accuracies are calculated according to the users need and top 5 books are displayed



```
Recommendations for Night Whispers:

1: Portrait in Death, with distance of 0.5236238521039133:
2: Macgregor Brides (Macgregors), with distance of 0.5521514539031239:
3: Small Town Girl, with distance of 0.5563508193381498:
4: Temptation, with distance of 0.5806862221664923:
5: Betrayal in Death, with distance of 0.582167748151904:
```

Figure 4.3 Related books with users need

5. CONCLUSION:

This system of recommendation is simple and convenient for user to use and reduce searching efforts, while using method by avoiding answers of complicated questions and other method depending on rating of data.

The complete work signifies that the similar users are based on the relevant distance. The similarity is based on the maximum size of cluster and relevant value.On the basis of price, quality, publisher and author, recommendation is done. Academician and students are more preferred by the book recommendation system.KNN is a machine learning algorithm to find clusters of similar users based on common book ratings, and make predictions using the average rating of top-k nearest neighbors.

REFERENCES

1.“A Survey of Recommender System Techniques, Challenges,” International Journal of Emerging Technology and Advanced Engineering,T. Zuva, vol. 2, November 2012.

2."Book recommendation system based on combine features of content based filtering, collaborative filtering and association rule mining.", A.S. Tewari, A. Kumar, and A.G. Barman,IEEE International. IEEE.

3. “Improved Collaborative filtering approach based on User similarity Combination”,IEEE International Conference on Management Science &Engineering .

4.“A Collaborative Approach for Web Personalized Recommendation System”,Nirav M. Khetra, ShrutiB.Yagnik, IJEDR