A

SEMINAR REPORT

On

**RECOMMENDATION SYSTEMS**

**USING**

**K-NEAREST NEIGHBOUR ALGORITHM**

Submitted in partial fulfillment of the requirements for the award of degree of

**BACHELOR OF TECHNOLOGY**

IN

**COMPUTER SCIENCE**

****

****

Submitted by:

**AAKRITI SINGH**

Smart Card ID: **BTBTC19194**

Exam Roll No: **1912902**

**DEPARTMENT OF COMPUTER SCIENCEE**

**BANASTHALI VIDYAPITH**

**July-Dec, 2021**

ABSTRACT

One of the potent personalization technologies powered the adaptive web is recommendation systems. It predicts user preferences for products and services by learning past user-item relationships from a group of user who share the same preferences and taste. It brings together the opinions of large interconnected communities on the web, supporting filtering of substantial quantities of data. In this report, I have introduced core concepts of recommendation system, its primary applications for the users of adaptive web, how similarity is computed in recommendation systems using K-nearest neighbor algorithm which states that, “Similar things exists in close proximity”, pros and cons along with future scope and conclusion.

CONTENTS

|  | |
| --- | --- |
|  | |
| 1. **Introduction** 2. **Methodology / algorithm / tools & technology** 3. **Applications** 4. **Benefits / Limitations** 5. **Conclusion** 6. **References** |  |

**INTRODUCTION**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

When users interact with large items present in the catalogue of E-commercial sites then there are majorly two ways of user-platform interaction:-

* **User knows what they are looking for**

In this case, user directly writes the query and the system gives the result.

Here, user has pulled the information from the system.

* **User actually don’t know for what they are looking for**

In this case, recommendation systems comes into action which actually

pushes the information towards the user.

Recommendation systems, recommends the user certain items and products based on past data of user and finds match between user and item filtering relevant information from large volume of data and taking care of user’s interests and preferences.

There are many techniques to implement recommendation systems and one among them is **collaborative filtering**. This technique filters or evaluates items using the opinions of other people. Using this technique, recommendation systems are able to provide an accurate prediction when enough data is provided because it is based on user’s preference.

It is also divided into two types:

* **User based collaborative filtering**

It is an effective way of recommending useful contents to users by exploiting the intuition that a user will likely prefer the items preferred by similar users. Therefore, algorithm at first tries to find the user’s neighbor based on user’s similarities and then combines the neighbor users rating score by supervised learning like k-nearest neighbor.

* **Item based collaborative filtering**

Item based collaborative filtering also works in the same way as user based collaborative filtering works in terms of user’s rating score. Instead of nearest neighbor,

it looks into set of items; the target user has already rated items and this algorithm computes how similar items are to the target item under recommendation. After that it also combines the customer’s previous preferences based on these item’s similarities.

On one hand, recommendation systems has revealed some potential challenges such as cold start problem and data scalability whereas on the other it has been effective in several domains.

**METHODOLOGY / ALGORITHM / TOOLS AND TECHNOLOGY**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Algorithm**  : K-Nearest Neighbor Algorithm

**Approach**  : Movie recommendation systems using Collaborative user based filtering

**Problem statement**-

1. Consider user X.

2. A group of user likes the movie as X does and dislikes the movie as X

does. These set of users are neighborhood of user X.

3. Then we find other movies liked by these set of users N and recommend it

to user X.

**Formal model of recommendation system**-

U: Set of users

S: Set of movies

p: utility function

U X S -> R

where R denotes rating given by user to movie.

Objective: Learn ‘p’ from the data

Output: When new unlabelled data is given it produces appropriate output

**Steps of the algorithm:**

1. Generate user-item two-dimensional matrix of rating as Rmxn , where each rating is ru,i.

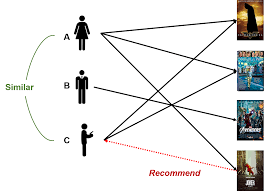
2. Use principle of Pearson correlation similarity to compute similarity between 2 users and generate user-similarity matrix.

3. According to result obtained in step 2, find K number of ratings which has maximum value and these corresponding K users forms the neighborhood of user X.

4. Compute the predictive value of i for target user X using the weighted average formula

rxi=∑ Sxy \* ryi

∑ Sxy

****

**Step 1:**

**Generate user-item two-dimensional matrix of rating as Rmxn , where each rating is ru,i.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ITEM**  **USER** | **Harry Potter1** | **Harry Potter2** | **Toy Story** | **StarWars** |
| **A** | **5** | **3** | **1** | **5** |
| **B** | **2** | **4** | **4** | **2** |
| **C** | **4** | **3** | **1** | **?** |
| **D** | **4** |  | **4** | **2** |
| **E** | **1** |  |  | **4** |

Load the dataset and generate 2-dimensional matrix of rating Rmxn such that,

m=Set of users who have reacted to the item

n= Set of movies in the system

The values in the matrix denotes rating on the scale 0 to 5. Assume C to be the target user to whom movie has to be recommended and let that movie be Starwars .

**Step 2:**

**Use principle of Pearson correlation similarity to compute similarity between 2 users and generate user-similarity matrix.**

2.1) Normalize ratings by subtracting row means.

2.1.1) Calculate row means for given row respectively.

Avg(A)=14/4 ; Avg(B)=12/4 ; Avg(c)=8/3

Avg(D)=10/3 ; Avg(E)=5/2

where Avg means average.

2.1.2)Subtracting row means from individual rows respectively except

missing values.

The modified rating matrix->

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Harry Potter1 | Harry Potter 2 | Toy Story | Starwars |
| A | 6/4 | -2/4 | -10/4 | 6/4 |
| B | -4/4 | 4/4 | 4/4 | -4/4 |
| C | 4/3 | 1/3 | -5/3 | ? |
| D | 2/3 |  | 2/3 | -4/3 |
| E | -3/2 |  |  | 3/2 |

2.2)

Consider row 1 of the modified rating matrix as user A’s rating vector and similarly for user B,C,D,E and F.

Compute Cosine similarity using cos Ө where Ө represents angle between two vectors.

cos(rC,rA)=0.837

cos(rC,rB)=0.617

cos(rC,rD)=-0.063

cos(rC,rE)=-0.436

**Step 3:**

**According to result obtained in step 2,find K number of ratings which has maximum value and these corresponding K users forms the neighborhood of user X.**

Since the problem is classification problem, select K value as odd because we have to return the mode value. Let K=3, so the top K users that forms the neighborhood are A, B and D.

**Step 4:**

**Compute the predictive value of i for target user X using weighted average formula**

**rxi=∑ Sxy \* ryi**

**∑ Sxy**

Substituting the values we get,

0.837\*6/4 + 0.617\*-4/4 + 0.063\*4/3

0.837 + 0.617 - 0.063

=0.51

**Since the predicted score of movie starwars is greater than 0 this means user C might also like this movie and hence the recommender system recommends the movie to user C**

**APPLICATIONS OF RECOMMENDATION SYSTEM**

**1. E-Commerce**

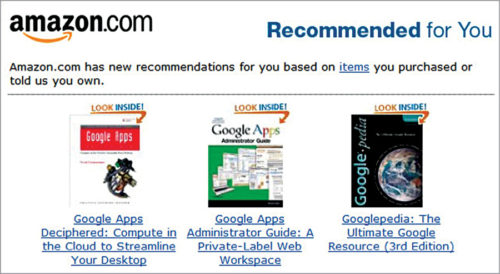
Recommender systems enhance e-commerce sales in the following ways-

* Browsers into buyers: Visitors to a website often look over the site without ever purchasing anything. Recommender systems can help customers find products

they wish to purchase.

* Cross-sell: It improves cross-sell by suggesting additional products for the customer

to purchase. If recommendations are good then average order size tends to increase.



**2. Media**

Similar to e-commerce, media businesses are one of the first to jump into recommendations.

It is difficult to see a news site without any recommendation.



3**. Banking**

Banking for masses are prime for recommendations. Knowing a customer’s detailed financial situation, along with their past preferences, coupled by data of thousands of similar users is

quite powerful.

**BENEFITS AND LIMITATIONS**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

* **BENEFITS**

1. **Drives traffic**:

A recommendation engine can bring traffic to the site which accomplishes this

with customized e-mails.

2**. Engage customers**:

Customers end up being more engaged in the website when individualized item recommendations are made.

3. **Boosts number of items per order**:

The number of products per order likewise increases with a recommendation.

When the customers is revealed options that fulfill his interest, he is most likely

to add choices to his purchase.

4**. Provides relevant content:**

By analyzing the customer’s present site use and his previous browsing history,

a recommendation engine can deliver appropriate product suggestions as he stores.

* **LIMITATIONS**

1. **Data sparsity:**

User-item input data matrix could have a few rating scores of the total number of items available, even though users are very active. In addition, because users tend not to rate actively, calculating similarity over set of items could be a challenge. These problems give rise to inaccurate performance of the recommendation system.

2. **Cold start problem:**

The cold-start problem is the difficulty of making recommendations when the users or the items are new. Here the recommendation system doesn’t have any past history, interests and preferences of the new user or item.

New User?

I don’t know the interests!

Lets create account on spotify!

**CONCLUSION**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Recommendation systems have been an important part of E-commerce sites on the web for

the customer to suggest items what they would be interested in. With the increase in number

of users and items, recommender systems encounter the major shortcoming: data sparsity

and cold start problem which bring out the reduced quality of prediction and it becomes

inefficient in time consuming. The alternative way is using item based collaborative

recommender systems that improve the prediction accuracy and recommendation quality.

To deal with cold start problem what we can do is that as the new user creates an account

on a platform, for example on Netflix then we could ask user to select top three genre that he

likes among the provided list with which we can take an idea to provide which type of movie

recommendations we should show to him.

We know the fact that, “Challenges were there, are there and

would continue to be there in the field of technology”, the thing matter is that how we

deduce the approach to resolve the limitations. After all recommender systems, by adding more

revenue to the business and creating large user engagements have proved to be one of the

greatest machine learning systems.

**REFERENCES**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**1.** A. R. Lahitani, A. E. Permanasari, and N. A. Setiawan, “Cosine similarity to determine similarity measure: Study case in online essay assessment,” In 2016 4th International Conference on Cyber and IT Service Management, Apr, 2016.

**2.** P. Phorasim and L. Yu, “Movies recommendation system using collaborative filtering and k-means,” International Journal of Advanced Computer Research

**3**. Balabanovíc, M., Shoham, Y.: Fab: Content-Based, Collaborative Recommendation. Communications of the ACM, (1997) 40(3).

**4.** https://towardsdatascience.com/intro-to-recommender-system-collaborative-filtering-64a238194a26

**5.** https://en.wikipedia.org/wiki/Collaborative\_filtering

**6.** https://builtin.com/data-science/collaborative-filtering-recommender-system