iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IX **Month of publication:** September 2022

DOI: https://doi.org/10.22214/ijraset.2022.46615

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IX Sep 2022- Available at www.ijraset.com

Recommendation System Using Matrix Factorization

Tejashree More[1], Prof. Surekha Kohle[2]
[1, 2]*Department of Master of Computer Application, Veermata Jijabai Technological Institute, Mumbai*

Abstract: ***In today's world, people are flooded with a lot of information, and no. of choices are overwhelming. For example, in any online shopping platform such as Amazon, if we search for a particular product, thousands of results appear and it becomes very difficult to select an item from vast pool of options. The growth of digital information and the number of users over the Internet has created a potential problem of information overload. The recommendation system solves this problem by searching through a large volume of data and providing personalized content to the user. This paper describes the introduction to the recommendation system, its three main types – content-based filtering, collaborative filtering, and hybrid filtering, and addresses the data sparsity problem. This paper proposed a collaborative filtering approach using matrix factorization to mitigate the sparsity problem and improve the quality of the recommendation.***
Keywords: Recommender Systems; sparsity problem; Matrix Factorization; Collaborative Filtering; Ratings

I. INTRODUCTION

The Recommendation System is one of the most popular applications of the machine learning system. It is an information filtering system that attempts to predict user's interests and recommend products that are likely to be of interest to users. Today lot of big companies like Amazon, Netflix, LinkedIn, Facebook, and YouTube are using recommender systems. For example, Amazon uses the recommendation system to predict the item that the user is most likely to buy based on past purchases, the rating that the user provides, their clicks, moves, and many other parameters. Similarly, Spotify uses a recommendation system to provide personalized suggestions to its subscriber about the songs, audio books or the podcasts that they would like to listen. The recommendation system predicts the rating that a user would give to an item and provides recommendations as per his interest. It speeds up the searches and allows the users to access the content that interests them in an easier and a faster way. From the user's point of view, it helps the user to get more personalized recommendations and from the company's point of view, it helps them to improve customer engagement on the website and enhances revenue for marketing, thus increasing profit.

II. TYPES OF RECOMMENDATION SYSTEM

There are three main types of recommendation systems that exist - Content Based Filtering, Collaborative Filtering, and Hybrid Filtering.

A. Content Based Filtering

The goal of content-based filtering is to classify products with specific keywords, learn what the customer likes, search for those terms in the data, and then make similar recommendations. It filters out a similar item based on the attributes of the item. For example, in the case of the movie, who is the actor of the movie, the director, the movie is of which genre, and so forth. For instance, if a user likes to watch action movies such as Spider-Man, then content- based filtering recommends similar movies with action genres such as Ant-Man, Iron Man, Avengers, etc. It attempts to guess what a user would like based on their previous actions and explicit feed- back. The drawback of content-based filtering is that it does not recommend items that are not searched by the user in the past.

Basically, it assumes that if a user was interested in a particular item in the past, they will once again be interested in the same thing in the future.

B. Collaborative Filtering

In collaborative filtering, recommendations are given based on other users who have similar preferences as that of the current user. It is based on collecting and analyzing data on users' behavior. For example, if user A likes Grapes, Cherry, and Orange while user B likes Mango, Grapes, and Cherry, they have similar interests. So, there are high chances that A would like Mango and B would like Orange. This is how collaborative filtering works. Collaborative filtering finds the users with similar interests, classifies them into clusters of similar types, and recommends items as per the preference of their cluster.

©IJRASET: All Rights are Reserved

C. Hybrid Filtering

The hybrid recommendation technique is a combination of content-based and collaborative filtering. It provides better recommendations as compared to using a single recommendation technique. Combining multiple recommendation techniques helps to overcome the limitations of using one particular technique.

III. RELATED WORK

During the last decade, a lot of work has been done on the development and improvement of recommendation systems. There are still a lot of unresolved problems in this area. Some of the work done to address the issues in the recommendations is as follows:

Kunal et al. [3] have portrayed the need for recommendation systems on multiple e-commerce platforms like Amazon, Flipkart and they have also discussed different types of recommendation systems that are available and the significant challenges related to recommendation systems such as cold start problem or the quality of the recommendations, scalability, and many more.

Jyoti et al. [5] have described an overview of recommendation systems. They have elaborated on different collaborative filtering techniques and have proposed a hybrid framework where they have combined item-based collaborative filtering with demographics-based user clusters in an adaptive weighted scheme. The proposed system addresses the cold-start problem and improves scalability. Paritosh et al. [5] have proposed a new system for more effective web page recommendation with the help of a hybrid filtering technique which is a combination of collaborative filtering techniques and pattern-finding algorithms. For better pattern discovery, the CHARM algorithm has been used. The main advantage of the CHARM algorithm is redundant rules can be eliminated, and because of that prediction time and quality is improved.

IV. PROBLEM STATEMENT

The recommendation system offers new ways to find or retrieve personalized information on the Internet. It enables users to easily access products and services within a limited period of time The general problem for recommender systems is that they often have very little data. Sometimes, there are users who have rated just a few items. If a user is not able to rate or like more items then somehow it is difficult process to determine his/her interest and it is also possible that user is related to wrong area. Sparsity problem occurs if very less information or data is available about user to give recommendation. If users do not give ratings or reviews to the items they have purchased or the content they have watched, the rating model becomes sparse because of which data sparsity problem occurs, it decreases the possibilities of finding a set of users with similar ratings or interests. This is an important problem as knowing whether or not it's worth pushing the users for more ratings and feedback is an important decision and knowing how well the new information improves the recommendations can be crucial.

V. PROPOSED RECOMMENDATION SYSTEM

In this paper, we try to solve the data sparsity problem by using the matrix factorization technique. The aim of this paper is to study how the matrix factorization technique helps to reduce the sparsity of the data and produce more accurate recommendations.

A. Matrix Factorization

Matrix Factorization is a collaborative filtering technique that identifies relationships between items and users' entities. Matrix factorization is a more effective technique as it helps us to discover the latent features by finding the underlying interaction between users and items. For example, in a movie recommendation system, there are a group of users and a set of items. Given that each user has rated some movies, we would like to predict the rating that a user might give to an item that has not been rated yet. Based on the ratings predicted by the matrix factorization, recommendations are given to the user.

In this case, all the ratings given by the users to the movies is represented as a matrix. Consider a situation where we have 4 users and 5 movies, ratings are the integer values ranging between 1 to 5, the matrix would look something like this,

TABLE I. USER'S RATING ON MOVIE

	Movie1	Movie2	Movie3	Movie4	Movie5
U1		2	4	5	1
U2	1			5	3
U3	1	1	2	3	1
U4	3		4	4	

Since not all users give ratings to all movies, there are a lot of missing values in the matrix, which results in a sparse matrix. Matrix factorization solves this problem by discovering latent features which determine how a user rates an item. For example, two users would give high ratings to a certain movie when the movie is directed by their favorite director or if the movie is a comedy movie, which is a genre preferred by both users. From the above table, we can find that U1 and U4 have given high ratings to Movie3 and Movie4". Thus, from matrix factorization, we are able to find the relationship between users and items, discover the latent features and predict a rating based on the similarity between users' preferences and interactions.

B. Mathematical Equation

Consider a set of users as Ui and a set of items as Vj. The goal is to find the association between the user and item by discovering k latent features and predicting the rating that a user would give to an item. In this technique, a bigger rating matrix is decomposed into two smaller lower dimensional matrices. The first matrix represents the row for each user and the second matrix represents the column for each item. The predicted rating for a user i with item j is computed as a dot product of the user feature matrix and item feature matrix.

Mathematical Equation is given by:

$$R = U_i V_j^T \quad (1)$$

Here, matrix U represents the user feature matrix, which is an association between users and features (m x k) while matrix V represents the item feature matrix, which is an association between items and features (k x n). Both of the matrices have k dimensions (i.e., k latent features).

To train the model, we minimize the difference between actual ratings and predicted ratings through iterations. This results in the following objective function of Matrix Factorization (MF),

$$min \sum_{i,j \in k} \left[L(U_i, V_j, R_{ij}) + \gamma_i ||U_i|| + \gamma_j ||V_j|| + + \lambda_i ||U_i||^2 + \lambda_j ||V_j||^2 \right] \quad (2)$$

Here, i, j is the index of i^{th} user and j^{th} item in the matrix R. Ui represents the individual row of User matrix and Vj represents an individual row of Item matrix.

L (U_i, V_j, R_{ij}) presents the loss function and is given by,

$$L = \sum_{i,j \in K} (R_{ij} - UV^T) \quad (3)$$

Here, K is a set of (i,j) pairs for which R_{ij} is known(i.e.,training set) and γ, λ are optional hyperparameters for regularization.

Here,

$$UV^T = \sum U_i \, . V_j \quad (4)$$

The system learns the model by fitting the previously observed ratings, that is training data. However, the main aim of this technique is to generalize those previous ratings in such a way thatit predicts the unknown ratings. Thus, the system must therefore avoid overfitting the training data by regularizing the learned parameters.

C. Example

Consider a movie recommendation system. Let us consider U1, U2, U3, U4, and U5 to be the users of the system and M1, M2, M3, M4, and M5 to be the movies of the system. The following matrix represents the ratings given by the users to the movies.

Table II. Initial rating matrix

Users\Movies	M1	M2	M3	M4	M5
U1	3	1	1	3	1
U2	1	2	4	1	3
U3	3	1	1	3	1
U4	4	3	5	4	4

Here, a big rating matrix is represented as dot product of two rectangular matrices with smaller dimensions (i.e., user matrix and item matrix)

Table III. Dot product of user feature matrix and item feature matrix

	M1	M3	M3	M4	M5
U1	3	1	1	3	1
U2	1	2	4	1	3
U3	3	1	1	3	1
U4	4	3	5	4	4

=

	F1	F2
U1	1	0
U2	0	1
U3	1	0
U4	1	1

.

	M1	M2	M3	M4	M5
F1	3	1	1	3	1
F2	1	2	4	1	3

Rating Matrix User Matrix Item Matrix

Here, one matrix is a User Feature matrix where rows represent users and columns represent k latent factors and another matrix is an Item Feature matrix where rows represent k latent factors and columns represent items. Matrix factorization finds the similarity between the preferences and interactions of the users. The goal is to find all the user-features dependencies in the matrix.

The task here is to learn the interaction between item and user such that their dot product represents their corresponding observed interaction. We start with some random values and we minimize the differences by performing multiple iterations until we get close to actual values. This technique is called gradient descent.

So, the values in the above matrices can be written as

	M1	M2	M3	M4	M5
U1	1.46	1.37	2.26	0.7	0.59
U2	1.32	1.53	1.85	0.91	0.5
U3	2.76	3.37	3.73	2.07	1.02
U4	1.68	1.99	2.32	1.2	0.63

=

	F1	F2
U1	0.2	0.5
U2	0.3	0.4
U3	0.7	0.8
U4	0.4	0.5

	M1	M2	M3	M4	M5
F1	1.3	3.1	0.3	2.5	0.2
F2	2.4	1.5	4.4	0.4	1.1

From the above representation, we can see that values in the rating matrix are a dot product of the user and item matrix.

For example, the value in the first cell (1.44) can be retrieved as

$$1.46 = (0.2 * 1.2) + (0.3*2.4)$$

If we compare 1.46 with the value in the original rating matrix, that is 3, then we need to increase the values in the user/item matrix in order to get a value closer to the original rating matrix.

In this way, we can repeat this process by changing the value until we get the value closer to the observed ratings. Using this technique, missing entries in the rating matrix are replaced with the calculated predicted ratings.

Thus, on the basis of predicted ratings, the system knows what to recommend to users with unseen movies.

VI. CONCLUSION

The recommendation system makes the job of the online user easy by presenting a series of products that could interest a user. The focus is to have good quality and accurate recommendations. But still, the recommender system faces some challenges like performance, quality of recommendations, the privacy of data, and many more. There are a lot of improvements that can be made in the future to overcome the challenges that exist in current recommendation systems.

This paper briefly describes about the recommendation system, and its types, and proposed a model based on matrix factorization to resolve the data sparsity problem.

REFERENCES

[1] Mansur, Farhin, Vibha Patel, and Mihir Patel. "A review on recommender systems." 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). IEEE, 2017.

[2] Vaidya, Nayana, and A. R. Khachane. "Recommender systems-the need of the ecommerce ERA." 2017 International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2017.

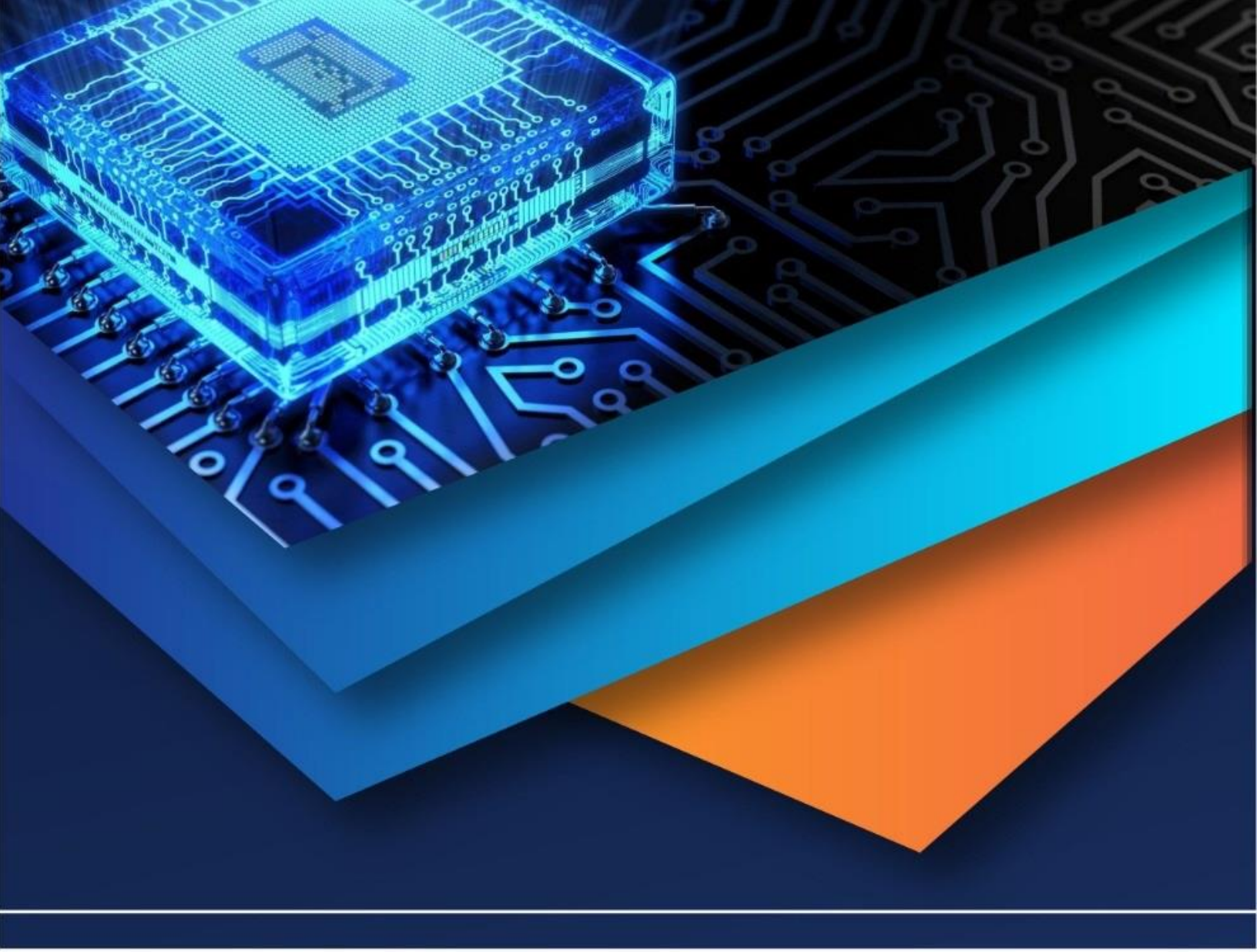
[3] Shah, Kunal, et al. "Recommender systems: An overview of different approaches to recommendations." 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). IEEE, 2017.

[4] Chen, Ming. "Research on recommender technology in E- commerce recommendation system." 2010 2nd international conference on education technology and computer. Vol. 4. IEEE, 2010.

[5] Gupta, Jyoti, and Jayant Gadge. "A framework for a recommendation system based on collaborative filtering and demographics." 2014 international conference on circuits, systems, communication and information technology applications (CSCITA). IEEE, 2014.

©IJRASET: All Rights are Reserved

[6] Nagarnaik, Paritosh, and A. Thomas. "Survey on recommendation system methods." 2015 2nd International Conference on Electronics and Communication Systems (ICECS). IEEE, 2015.
[7] http://www.quuxlabs.com/blog/2010/09/matrix- factorization-a-simple-tutorial-and-implementation-in-python/
[8] https://towardsdatascience.com/recommendation-system- matrix-factorization-d61978660b4b
[9] https://www.analyticsvidhya.com/blog/2021/07/recommen dation-system-understanding-the-basic-concept



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)