

Movie Recommendation System Using SVD

†Bhagyashree Shelke^[1], Suraj Mandhane^[2], Rishav Agrawal^[3], Prasanna Jain^[4], Umakant Mandawkar^[5]

^[1, 2, 3, 4]Students of CSE Dept., Sandip University, Nashik (MS), India

^[5]Asst. Professor CSE Dept., Sandip University, Nashik (MS), India

Abstract-

Movie recommendation is worldwide known concept. The primary goal of any recommendation system is to recommend an applicable movies and data to the consumer. There are lots of parameter in this scenario which we consider for recommendation filtering of our objects. The SVD is collaborative filtering technique which is work as user based as well as item based also. For better movie recommendation there are many algorithms and systems which are available in market. For the best results SVD is used. SVD is Singular Value Decomposition technique which is widely used in machine learning algorithms. It requires the knowledge of matrix algebra as well as vector calculus. This technique provides fast, easy and accurate recommendation system. Intelligence researchers noted the test result of SVD which is correlated with human intelligence. It is also known as “Factor Analysis”. The big platform like Netflix, Amazon Prime, Hotstar and Pandora use SVD in their form for better result. The algorithm and technique of SVD is short and easy. SVD comes with many interesting algebraic properties. It mines the user required or item based data very correctly than other models. SVD works with different languages but Python is more efficient when it comes to the machine learning or deep learning project. Software like Anaconda with python having some imported libraries used in SVD for the better analysis and results. So, for better result we focused on SVD.

Keywords- SVD, Vectors, Recommendation, Collaborative filtering, etc.

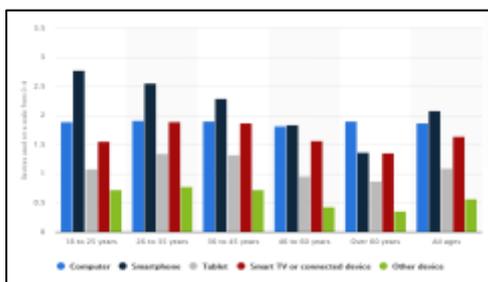
1. INTRODUCTION:

On this planet each and every human being have different choices regarding everything. Same goes for entertainment field. Movies are the biggest source of entertainment in 21st century. But it's not necessary that all kind of movie liked by everybody. Everybody have their personal taste and they prefer only those content for their entertainment. Recommendation systems have become very popular in the field of machine learning, and are a huge advantage for technology giants such as Netflix and Amazon in terms of targeting their content to specific audiences. There are different methods and algorithms through which we can perform recommendation of movies. Collaborative filtering algorithms are one of the oldest algorithm or technique in movie recommendation system. But it having many problems like cold start problem. So we come up with the solution to combine two techniques collaborative filtering and SVD. SVD means Singular Value Decomposition is one of them. SVD is well used tool for recommendation system throughout the world. It is general purpose numerical tool in algebra for data processing. It is basically data reduction tool in which we have mainly high dimensional data like big dataset or big images data. So SVD helps us to reduce that data into key features for analyzing, understanding and describing the huge amount of data. It is

data driven generalization of the Fourier transform (FFT). The SVD is going to allow us to do is to tailor a coordinate system or a transformation based on the data that we have. SVD used to solve matrix sequence of equation

$$Ax = B$$

Linear system of equation for non-square 'A' matrix and we particularly used it for linear regression models [1]. SVD used in Facebook, Google, Microsoft and many more. The worldwide famous recommender system like Amazon, Netflix use the SVD where it basically builds a correlation pattern of people's recommendation. Watching movies on Television is an old concept now because when you get the entertainment which you prefer then televisions are outdated.



[12]

Fig. 1 General Percentage of age group of people and devices they used.

2. LITERATURE REVIEW:

Bei Bei Cui paper on movie recommendation system which is based on KNN model gives that they used KNN (K nearest neighbor) the filtering algorithm is combined with the algorithm. Then they explained the principle and architecture of the relational database model of the JAVAEE system. They tried personalized recommendation system and gave preference to collecting and analyzing historical behavior of person. Personalized recommendation system is a kind of information filtering technology. It is a comprehensive system that combines various data mining algorithms and user information. [2]

Bhumika Bhatt, Prof. Premal J Patel, Prof. Hetal Gaudani gives Recommendation system Divided into three categories: collaborative filtering, content methods and hybrids. In this document, collaborative filtering is divided into two types: memory-based and model-based recommendations, and their applications and limitations. Demand from customers throw new challenges like overloading on system and hence the recommendation system still needs to be improved to become a better system. The three methods suggested by the researchers. Collaborative filtering system, content-based system, hybrid system. Provide new methods that take quality and confidentiality into consideration, and can make recommendations in a wide range of applications. Therefore, the current recommendation system needs to be improved according to current and future needs to improve the quality of recommendations. [3]. Steven Postmus, used some real time data analysis in this paper to show popularity and increasing business in this field and grows their business from one video shop to the biggest business in entertainment in 2 decades. In this study, we will initially only use customer ratings, and then supplement them with external metadata such as actors, genres, IMDb ratings, and release date. This work also uses collaborative filtering, content-based filtering, and hybrid filtering. But the problem as the weakness of collaboration and content filtering is the problem of dealing with new users or projects. [4] S. Rudraksha, V. Munot, S. Mishra, S. Misal, Y. Kandhare, A. Kulkarni used content-based filtering, collaborative filtering and Hybrid filtering by using the above two methods. In hybridization, there are different approaches like weighted hybridization, switching hybridization, mixed hybridization, feature combination, cascade [5]. SRS Reddy, S. Nalluri, S. Kuniseti, S. Ashok and B. Venkatesh used content based movie recommendation system such kind

Recommendation systems are useful for organizations that collect data from a large number of customers and want to effectively bid for the best. The content-based filtering system is divided into three methods: packaging method, filtering method and built-in method. [6]. Scharf & Alley The author proposes a flexible multi-component dosage recommendation system to predict the optimal fertilization dosage for winter wheat. [7]. Von Reischach et al. The author proposes a ranking concept that enables users to generate ranking criteria. [8]. Goldberg et al. The author proposes a joint filtering method [9]. Mojdeh Saadati, Syed Shihab, and Mohammed Shaiqur Rahman provide different types of recommendation systems, mainly a comparative study of recommendation systems that SVD gives the most accurate results. [10].

“Table 1 Summary of all models with test and train RMSE value” [11]

S.N O	MODEL	TRAIN RMSE	TEST RMS E
0	XGBOOST	0.810569	1.072 28
1	BASELINEON LY	0.881143	1.067 84
2	XGB_BSL	0.8 021 1 2	1.068 88
3	KNNBASELIN E_USER	0.304498	1.067 65
4	KNNBASELIN E_ ITEM	0.181651	1.067 65
5	XGB_BSL_K NN	0.810123	1.071 74
6	SVD	0.891529	1.067 66
7	SVDPP	0.787158	1.067 5
8	XGB_BSL_K NN_MF	0.810568	1.068 7
9	XGB_KNN_M F	1.07269	1.072 76

3. EXISTING SYSTEMS:

There are several systems which are implemented in this field. But they are still improving. The reason for this improvement is the popularity of organizations like Netflix that focus on customer satisfaction. Before the advent of recommendation systems, people would physically select movies from the movie library to watch. They must read user reviews and choose a movie or random movie based on the reviews. This method is impractical because many viewers have unique preferences for movies. In the past ten years, many recommendation systems with different methods have been developed, such as collaborative methods, content methods, hybrid methods, etc. By studying the behavior and history of various customers based on feedback, the system will suggest to you who you are. These recommendation systems are divided into two types: collaborative filtering technology and content-based applications. The collective method combines comments from different users with similar tastes and then recommends movies, while the content-based method is limited to one user and uses past ratings and user history to make recommendations. Various techniques have been proposed to implement the proposed structure, covering several areas of data mining, clustering, and red Bayesian techniques. The results got through this multiple models are always different and their accuracy changes. SVD is also implemented but there are a lot of scope there to get new technique in that. So, we tried that in this paper. Let's see the methodology.

4. DESCRIPTION OF THE RESEARCH WORK:

In the proposed model we used SVD with collaborative filtering system in consideration of user based as well as item based also. The parameters which we consider here are movie name, movie genre and movie rating. We used this parameter because generally young

customers prefer the genre and ratings of the movie. Getting this parameter gives us the most significant results as compared to other parameters. These parameters can vary as per service provider requirements. The tools used are Anaconda Jupiter with Python and some imported libraries. The concept of SVD of linear regression and RMSE is used. If any customer search for the movie, then that person will type the name and after that the number of recommendation they wanted like 5 or 10. It will show the result according to it. Though it is not a web application but by giving the front end link we can have proposed the web application also.

5. METHODOLOGY:

In this section we have briefed the methodology which we have used while implementing this method.

A. DATASET:

We have used a MovieLens 100k dataset in which contains lots of parameters in consideration. This dataset we take online. This data was collected from some time at MovieLens website. 100,000 ratings from 1000 users on 1700 movies. This dataset contains 943 users on 1682 movies. For this dataset we have to load some packages which are u.data, u.info, u.item, u.genre. By merging Movie_dataset and Dataset on the basis of Movie_id the new merged dataset formed called merged_dataset. In which contains user_id, movie_id, ratings, movie_name. The purpose of merging this dataset is avoiding duplicate data entries so they will merge into one so data processing will be easy.

B. DATA VISUALIZATION:

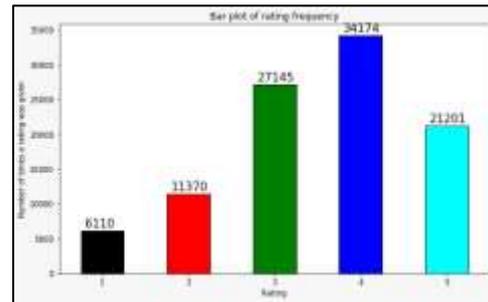


Fig. 2 Ratings based visualization

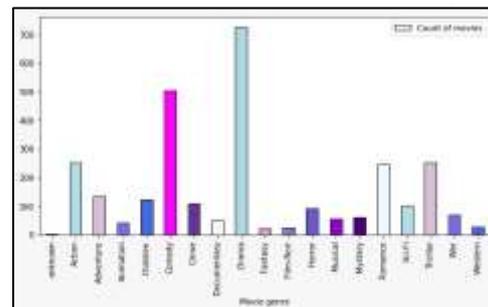


Fig. 3 Genre based visualization

C. IMPLEMENTATION:

I. SVD:

SVD is simply done on utility matrix and latent features of rows and columns. Formula is,

$$A = USV^T$$

Where, A is (m*n) utility matrix, U is (m*r) orthogonal left singular matrix, V is (r*n) orthogonal right singular matrix, S is (r*r) diagonal matrix. U indicates relationship between users and latent factors, V indicates similarity between items and latent vectors, and S represents strength of each latent factor. SVD is clearly a represent the relationship between users and movies. A function is defined to calculate the cosine similarity in a given data frame, and numpy one is used to extract the requested number of closely matched movie indexes. This function evaluates Einstein's convention for summing the operands. Interactive user interface.

II. COLLABORATIVE FILTERING:

It is a first recommender system which is produced when it comes to movie recommendation system in which to aggregate the evaluation. In this model each user is represented as vector which size is N where, N is the identical value which is offered by the business. N is a very large number so generally it is sparse. With rows and columns this all vectors put in the matrix which represents users and items respectively. By the result of that matrix we get utility matrix or rating matrix of size (Nu * Ni). Only using of collaborative matrix alone always lead us to the cold start problem so, the accuracy gets reduced certainly. So, we tried and refer the SVD and Collaborative filtering together. Collaborative filtering has item based and user based recommendations system and we both used in this model. Item based used for ratings and used based used for genre and movie.

III. IMPLEMENTATION AND RESULTS:

After implementing this model when user enter any movie name the recommendation system will recommend the movie with their released year. It gives you as much as recommendation you want. If somebody wants 5 then it will give 5 recommendations except the movie which we are using for recommendation. If webpage is made of this implementation, then user will be satisfied by results of this system. We got the results like –

```

Enter the Movie name: Toy Story (1995)
How Many recommendation You Want: 10
Top 10 Recommendations of- Toy Story (1995) are:
-----
16 - Apollo 13 (1995)
543 - Beauty and the Beast (1991)
753 - Amityville: Dollhouse (1996)
281 - Emma (1996)
8 - Aladdin (1992)
114 - Haunted World of Edward D. Wood Jr., The (1995)
596 - Hunchback of Notre Dame, The (1996)
450 - Transformers: The Movie, The (1986)
166 - Mystery Science Theater 3000: The Movie (1996)
923 - Island of Dr. Moreau, The (1996)
    
```

Fig. 5 1st result of movie Toy Story with 10 recommendations

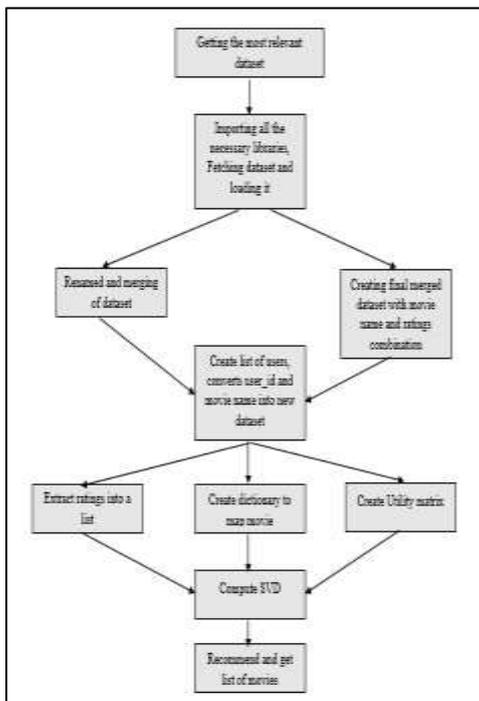


Fig. 4 This is the block diagram of our project implementation. The given blocks refer the method and steps which we used at the time of implementation.

```

Enter the Movie name: 101 Dalmatians (1996)
Enter Number of movie recommendations needed: 10
Top 10 movies which are very much similar to the Movie- 101 Dalmatians (1996) are:
Black Beauty (1994)
Free Willy 2: The Adventure Home (1995)
Evening Star, The (1996)
Robin Hood: Men in Tights (1993)
Cool Runnings (1993)
Turbo: A Power Rangers Movie (1997)
Remains of the Day, The (1993)
City Hall (1996)
Children of the Corn: The Gathering (1996)
    
```

Fig. 6 2nd result of movie Mrs. Dalloway with 10 recommendation

6. CONCLUSION:

By comparing SVD model with another models of different research paper we got more perfect results in it. The RMSE is low of the SVD model means this model is more accurate than others. SVD use collaborative filtering with item based and user based modelling for avoiding the issues like cold start problem. If the axis of normalization

of utility matrix get changed then a better recommendation system can be build. As in result this project is able to let you decide how many recommendations you want. It considers the different parameters at the same time like movie id, movie name, ratings, year, and number of recommendation. Ex. If you suggest Toy story so, the genre of Toy story is animation, comedy and fictional so it will not only show animation but combinely it will give you a result which is more likely near to the Toy story genre movie like Aladdin which is animated but Apollo 13 which is fictional as shown in result.

7. FUTURE SCOPE:

- a) Can develop a web application
- b) Can Include the content based filtering for top most results.
- c) Can work on real time data with supercomputers.

8. REFERENCES:

- 1) <https://www.youtube.com/watch?v=gXbThCXjZFM>
- 2) Bei Bei Cui, 2017, 'Design and Implementation of Movie Recommendation System Based on KNN Collaborative Filtering Algorithm', ITM Web of Conferences, 04008, itmconf/20171204008, pp. 800-805.
- 3) Bhumika Bhatt, Prof. Premal J Patel, Prof. Hetal Gaudani, 2014, 'A review paper on Machine learning based recommendation system', 2014 IJEDR, Volume 2 Issue 4, ISSN: 2321-9939, 2014, pp. 3995-3961.
- 4) Steven Postmus, 2018, 'Recommender system techniques applied to Netflix data', Virje University Amsterdam, February 2018.
- 5) Shardul Rudraksha, Vipul Munot, Satyadev Mishra, Swapnil Misal, Yogita Kandhare , Abhilasha Kulkarni, 2019, 'Hybrid approach to build anime recommendation system', 2019 IJRSET, Volume 8, Issue 11. November 2019, pp. 8362-8365.
- 6) SRS Reddy, Sravani Nalluri, Subramanyam Kunisetti, S. Ashok and B. Venkatesh, 2018, 'Content-based movie recommendation system using genre correlation', VIT University, Conference paper, November 2018, pp. 391-397.
- 7) Scharf, P. C., & Alley, M. M. (1993). Spring nitrogen on winter wheat: II. A flexible multicomponent rate recommendation system. *Agronomy journal*, pp. 85(6), 1186-1192.
- 8) Von Reischach, F., Guinard, D., Michahelles, F., & Fleisch, E. (2009, March). A mobile product recommendation system interacting with tagged products. *2009 IEEE international conference on pervasive computing and communications* pp. 1-6. IEEE.
- 9) Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), pp. 61-70.
- 10) Mojdeh Saadati, Syed Shihab, Mohammed Shaiqur Rahman Movie 'Recommender Systems: Implementation and Performace Evaluation', Iowa State University, Sep 2019.
- 11) M. Chenna Keshava, S. Srinivasulu, P. Narendra Reddy, B. Dinesh Naik, 2020, 'Machine learning model for movie recommendation system' April- 2020 IJERT, Volume 9 Issue 4, ISSN: 2278-0181, pp. 800-805.
- 12) <https://www.statista.com/statistics/84374/online-video-devices-by-age>

