

Movie Recommendation System using Simple Recommender-Based Approach

Sushmita Roy

(Author), Post-Graduation Student

Department of Information Technology (IT)

Thakur College of Science and Commerce (TCSC)

Mumbai, India

Abstract:- With the expeditious development of Information and Communications Technology and Web technologies, an enormous information is currently available and this results in a condition which in some cases is known as information overload. Due to these circumstances, it's growing arduous for a person to discover and to access information for taking decisions expeditiously to arrive at an effective conclusion. To perorate this nut, there are filtering systems for information, such as the recommendation system or recommendation engine, considered here in this paper, that help a person in identifying significant, tectonic and possible services or products of interest based on the preferences given by him/her. Several approaches exist like - simple recommendation, popularity-based recommendation, collaborative filtering, content-based filtering demographic-based filtering and keyword or metadata-based filtering. This study has been undertaken to investigate the simple recommender-based approach. The efficiency of this proposed methodology is verified by experiments based on The Movies Dataset.

Keywords:- Recommendation System; Recommendation Engine; Movie Recommendation System.

I. INTRODUCTION

Quite often we might think which algorithm IT giant like Google would be using to maximize its target ad revenue or often we wonder how e-commerce websites lure us to buy products through their choices such as “people who have bought product X also have bought Product Y” or “customers who have bought this have also bought this” or how can Facebook suggest us friends or suggest to tag friend in a group of pictures.

The answer to all of these is using a recommender system or recommendation engine. Now, with the large amount of information available on the Web and with the significant increase in the number of people, it is growing important for organizations to search and provide with the relevant information according to their preferences and tastes.

Currently organizations are involved into building recommendation systems or recommendation engine by learning and analyzing the past behavior of the customers. Hence, giving users with recommendations and choices of

their interest in instances like - Job posts, Online shopping, Movie of Interest, Suggested Video, Games of Interest, Facebook friend that you know and People who have bought this have also bought this etc. have now become easier.

II. WHAT IS RECOMMENDATION SYSTEM OR RECOMMENDATION ENGINE ?

Recommendation Systems, often termed as Recommender Systems or Recommendation Engines, are simply the usage of various algorithms whose aim is to provide relevant and accurate products to customers by means of filtering of useful data from a large pool of information base. Recommendation systems discover patterns in the data set by learning and analyzing user choices and behavior and produce the outcomes that best relates to the interests of user.

Thus, recommendation systems predicts the “preference” or “rating” that a customer would give to a product.

Figure 1 shows a typical usage of recommendation system.



Fig 1:- Recommendation System Usage

III. RECOMMENDATION SYSTEMS – THE NEED OF THE COMMERCE ERA

Recommendation systems generally belong to the group of Information Retrieval combined with some Data Mining and Machine Learning concepts. It is a branch or domain under the Data Science field and though progress has been achieved in this field, still it is growing area of research. These recommender systems are now of vital importance in the e-commerce industry. Recommender system recommend items or products to people such as books, games, movie/video, electronic product, online shopping products and other classes of products. Recommender system help the customers to get recommendations that are personalized, additionally helping users to make apt decisions in the online purchases, rise in the sales and redefining the users browsing experience, retaining the customer base and enhancing their shopping experience. Often, the “information overload problem” is being solved by these search engines, but they do not aim to provide data personalization. Recommendation engine thus help a step ahead by providing personalization.

A recommender system is thus a major technique that handles the “information overload problem” of Information Retrieval by advising its customers with apt and relevant items. Today, many recommendation systems are being built for the domains differently. But the ones developed are not accurate to satisfy the information needs of the customers. Hence, it is obviously necessary to develop or build good, effective and high quality recommendation system. In designing such recommendation systems, designers often face issues and variety of challenges that needs to be addressed with proper attention.

IV. MOTIVATION

In order to analyze and understand the behavior and the tastes of the users, many organizations currently include recommendation system or recommendation engine in their websites in order to reach some business goals. Among these goals we have:

- the rise of traffic to the web sites,
- effective marketing policy elaboration tailored for meeting the customer tastes,
- or in general the promotion or advertisement of product.

V. APPLICATION FIELDS FOR RECOMMENDATION SYSTEMS

Some of the most widely used areas of recommendation systems include: -

- Books
- Documents
- News
- Research Articles
- Images
- Movies
- Search Queries

- Music
- Social Tags
- Shopping
- TV Programs, etc.

VI. OBJECTIVES FOR THE IMPLEMENTATION OF A RECOMMENDATION ENGINE

There can be several reasons for which businesses today use recommendation system: -

- To results in an increase in the quantity of products used by user: In cases of online shopping, this could be executed by “increase the quantity of item or product selling”, whereas in the case of news portal it could be translated to “increase in the number/quantity of news items being read by the user”.
- Allow the customer to select item or product that could be otherwise difficult to search or locate without an accurate and precise recommendation system: This functionality allows less-known products and the customers generally do not tend to use to distinguish themselves from the way the user uses, so she can notice them.
- Increasing the user satisfaction or customer satisfaction levels: A well-developed and built recommendation system lightens the knowledge of the customer with the website. Users will generally find the recommendation as interesting, relevant, accurate, precise and with a correctly designed human-computer interaction. It will be one which the user can use easily and will also enjoy using with the passage of time. This combination provided with precise predictions and rich user interface will increase the subjective evaluation of these systems. These in return will then increase the system usage and the probability that the recommendations will be accepted. This will also tend to protect the user loyalty.
- Help in understanding the user taste in a better manner: From the users’ choices and preferences which has been collected, service providers, developers, business managers, etc. could need to develop new marketing policy that would suit individual tastes.

VII. IMPLEMENTATION FRAMEWORK

The proposed system is being built using Python language and Django framework.

Python is an interpreted, object oriented and high level programming language which provides support for different module and package. This encourages the concepts of program modularity and code reuse.

Django is a free, open-source and a widely-used Python web application framework. Generally, frameworks are used to save developers from having to reinvent the wheel and to help reduce some of the need and overhead to start developing a new website right from the scratch. This common functionality which are used for building the web applications comes with the framework instead of using separate libraries.

➤ *Django MVT*

MVT stands for Model-View-Template. It is a software design pattern being used in Django. It has a collection of three vital components – Models, Views and Templates.

- Model – It helps to take care of the database. It is a data access layer that is used for data handling.
- Template – It is presentation layer which handles the User Interface (UI) fully.
- View – It is used to write the logic for main business. The view also interacts with the model to carry the required data and then renders it to the template.

Figure 2 shows the MVT-based flow as - User first sends a request for a resource to application. Now, Django is working as a controller. So, it checks if the resource is available in the URL. If the URL can map the resource correctly, then view is being called that interacts with the model and the template. This template is then rendered to the user. Django responds to the user and send this template as a response.

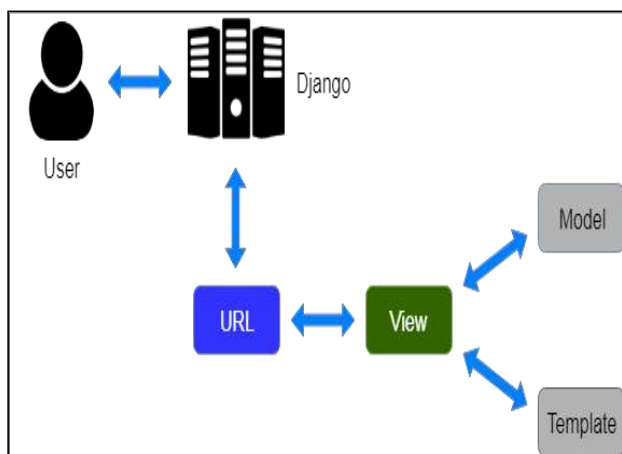


Fig 2:- MVT-based control flow in Django

VIII. RESEARCH METHODOLOGY

A. Research Goals

The goal of the Movie Recommendation System is to recommend movies to users which they have not watched yet. In achieving this goal, following two strategic objectives can be addressed: -

- Objective 1: To identify areas where relevant and accurate predictions could make the most difference.
- Objective 2: To assist users by recommending movies.

B. Source of Data Collection

For this study secondary data has been collected. the data relating to The Movies Dataset for the movies released on or before July 2017. These files contain metadata i.e. detailed data about the main data for all the thousands of movies listed in the MovieLens Dataset.

C. Simple Recommender System

The Simple Recommender offers generic recommendations to users based on the popularity of the movie and also based on the genre. The idea involved in this is - movies which are popular will have a higher probability of likes by audience.

All we have to do is sort our movies based on ratings and popularity and display the top movies of our list. As an added step, we can pass in a genre argument to get the top movies of a particular genre.

We can use IMDB's weighted rating formula to construct my chart. It can be represented mathematically as follows:

$$\text{Weighted Rating} = ((v/(v+m)) * R) + ((m/(v+m)) * C)$$

where,

- v – total number of votes accumulated for the movie
- m – minimum number of votes required for the movie to be listed in the charts
- R – average rating of the movie
- C – mean vote for the report

D. Approach

In the process of building a recommendation engine, we have implemented Simple Recommender-based approach with the intention of analyzing the dataset under consideration.

For executing the results using Python, first we need to import the required library files and load the dataset into a pandas data frame. Then we can retrieve the genre for all data records and if vote count exists. If it exists, we can retrieve the required values. Similarly, we can check for vote average. Then we can compute the mean of the above vote averages obtained and calculate the min number of votes that is necessary for the movie to be in the chart. Then we can filter the required movies that are qualifying for the charts and return the data. A function needs to be defined to calculate the weighted rating for each movie by defining a feature known as weighted rating and calculating it. Thus, we can then sort the data set based on this weighted rating.

Additionally, we can specify a genre and the data will be displayed based on that particular genre only to the user.

IX. RESULTS

Simple Recommender based approach: This system used overall TMDB Vote Count and Vote Averages to build Top Movies Charts, in general and for a specific genre. The IMDB Weighted Rating System was used to calculate ratings on which the sorting was finally performed.

Thus if I say, `print(pass results.head(15))`, the top 15 movies are displayed to the user as shown in Figure 3.

	title	year	vote_count	vote_average	popularity	genres	wr
15488	Inception	2010	14875	8	23.1081	[Action, Thriller, Science Fiction, Mystery, A...	7.917588
12481	The Dark Knight	2008	12369	8	123.157	[Drama, Action, Crime, Thriller]	7.916471
22079	Interstellar	2014	11117	8	32.2155	[Adventure, Drama, Science Fiction]	7.897387
2816	Fight Club	1999	9571	8	63.8686	[Drama]	7.892753
4863	The Lord of the Rings: The Fellowship of the Ring	2001	8882	8	32.8707	[Adventure, Fantasy, Action]	7.891187
252	Pulp Fiction	1994	8579	8	149.35	[Thriller, Crime]	7.886660
114	The Shawshank Redemption	1994	8351	8	51.5654	[Drama, Crime]	7.864880
7000	The Lord of the Rings: The Return of the King	2003	8216	8	29.3294	[Adventure, Fantasy, Action]	7.861927
231	Forrest Gump	1994	8147	8	48.3072	[Comedy, Drama, Romance]	7.856551
3014	The Lord of the Rings: The Two Towers	2002	7541	8	23.4255	[Adventure, Fantasy, Action]	7.851221
126	Star Wars	1977	5771	8	41.1497	[Adventure, Action, Science Fiction]	7.841185
1225	Back to the Future	1985	6211	8	25.7735	[Adventure, Comedy, Science Fiction, Family]	7.839812
134	The Godfather	1972	5824	8	41.1093	[Drama, Crime]	7.834947
1154	The Empire Strikes Back	1980	5981	8	13.471	[Adventure, Action, Science Fiction]	7.814881
44	Se7en	1995	5815	8	13.4574	[Crime, Mystery, Thriller]	7.811559

Fig 3:- Results displaying the top 15 movies

Thus, the top movie according to the charts is “Inception” which has the highest weighted rating as calculated by the system – 7.917588 and “Se7en” which has the weighted rating of 7.811669 appears to be the one with the least weighted rating and is basically at the last in the charts.

Now, if I say, `print(chart_genre('Romance').head(15))`, the top 15 movies are displayed to the user as shown in Figure 4.

	title	year	vote_count	vote_average	popularity	wr
10309	Dilwale Dulhania Le Jayenge	1995	661	9	34.457	8.565285
351	Forrest Gump	1994	8147	8	48.3072	7.971357
876	Vertigo	1958	1162	8	18.2082	7.811667
48251	Your Name.	2016	1030	8	34.461252	7.789489
883	Some Like It Hot	1959	835	8	11.8451	7.745154
1132	Cinema Paradiso	1988	834	8	14.177	7.744878
19901	Paperman	2012	734	8	7.19863	7.713951
37863	Sing Street	2016	669	8	10.672862	7.689483
882	The Apartment	1960	498	8	11.9943	7.599317
38718	The Handmaiden	2016	453	8	16.727405	7.566166
3189	City Lights	1931	444	8	10.8915	7.558867
24886	The Way He Looks	2014	262	8	5.71127	7.331363
45437	In a Heartbeat	2017	146	8	20.82178	7.003959
1639	Titanic	1997	7770	7	26.8891	6.981546
19731	Silver Linings Playbook	2012	4840	7	14.4881	6.970581

Fig 4:- Results when value is passed for a specific genre

Thus, the top romance movie according to the charts is “Dilwale Dulhania Le Jayenge” which has the highest weighted rating as calculated by the system – 8.565285 and “Silver Linings Playbook” which has the weighted rating of 6.970581 appears to be the one with the least weighted rating and is basically at the last in the charts.

X. CONCLUSION

This paper presents a movie recommendation system using Simple Recommender-based approach using Python. It inputs the data source for the movies Dataset for movies data released before July 2017 and provides recommendations for movies.

ACKNOWLEDGMENT

I would like to express my gratitude to my college, Thakur College of Science and Commerce (TCSC) for giving me this opportunity to accomplish the project work.

I express my heartfelt gratitude to the Principal Dr. (Mrs.) C.T. Chakraborty for her kind, inspiring and illustrious guidance and ample encouragement.

I am very thankful to the IT Department and Dr. Santosh Kumar Sing, Head of Department of Information Technology for his kind co-operation in the completion of this project.

Special thanks to my project guides, Dr. (Mr.) Santosh Kumar Singh and Prof. Mahendra Sharma who have given me the opportunity to do this, which even helped in doing Research and I learnt many new things.

REFERENCES

- [1] Sidnooma Christian Kabore. S 2012. Design and implementation of a recommender system as a module for Liferay portal.
- [2] N. Vaidya and A. R. Khachane. Recommender systems-the need of the e-commerce ERA. 2017 International Conference on Computing Methodologies and Communication (ICCMC), Erode, pp. 100-104. doi: 10.1109/ICCMC.2017.8282616
- [3] Khuro, Shah & Ali, Zafar & Ullah, Irfan. (2016). Recommender Systems: Issues, Challenges, and Research Opportunities. 10.1007/978-981-10-0557-2_112.
- [4] Venkatesan M. (Government Arts College (Autonomous), India) and Thangadurai K. (Government Arts College (Autonomous), India). History and Overview of the Recommender Systems. 2017. DOI: 10.4018/978-1-5225-0489-4.ch004
- [5] Francesco Ricci, Lior Rokach, and Bracha Shapira. Recommender Systems: Introduction and Challenges. Springer Science+Business Media New York 2015. DOI 10.1007/978-1-4899-7637-6_1
- [6] <https://www.analyticsvidhya.com/blog/2015/10/recommendation-engines/>
- [7] https://en.wikipedia.org/wiki/Recommender_system
- [8] <https://www.javatpoint.com/django-mvt>
- [9] www.kaggle.com
- [10] www.datacamp.com