IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 8, May 2023

Movie Recommendation System with Content Based Technique

Saurabh Gautre[1], Sanskar Tannirwar[2], Pratik Gupta[3], Anup Gain[4], Prof. Devashri Kodgire[5]

Students, Department of Information Technology[1,2,3,4]

Professor, Department of Information Technology[5]

Rajiv Gandhi College of Engineering Research and Technology, Chandrapur, Maharashtra, India

Abstract: *A recommendation system is a system that, depending on certain data, makes suggestions to users for specific resources like books, movies, songs, etc. The characteristics of previously loved movies are typically used by movie recommendation systems to anticipat ewhat movies a user would like. Such recommendation systems are advantageous for businesses that gather data from a lot of clients and want to successfully offer the finest recommendations. When creating a movie recommendation system, several variables may be taken into account, including the movie's genre, cast, and even director. The algorithms are capable of recommending movies based on a single attribute or a combination of two or more. The recommendation algorithm in this study is based on the kinds of genres that the user would want to watch. The method used to do this is content-based filtering with genre relevance. Movie Lens set of data is the one processed by the system*

Keywords: Movie Recommendations, Content-based Filtering, Cosine similarity.

I. INTRODUCTION

In recent years, improving the effectiveness of the commercial web services especially to use the personalized recommendation technology to realize the electronic commerce personalized service has gradually become a hot topic can cause widespread interest. But at present domestic most of e- commerce recommendation is usually: recommended best-selling products. Recommend related product according to user's browsing history is recommended so to speak, the first two recommended due to fundamental not considering the personality traits of the different users, therefore recommend simply does not have the characteristics of individuation, the third recommend a personalized composition, but most of the site also only stay in only the users against a person's purchase history, just for each user set up a personal purchase records, no transverse to the comprehensive information, so there is no collaboration recommended value which also is unable to realizes real-time comprehensive recommended goods.

II. METHODOLOGY

The procedure for gathering information for the literature review is described in this section. Information regarding movie recommender systems was gathered from sources that have undergone peer review. EBSCO Academic Search Premier, ScienceDirect, IEEE Library, ResearchGate, SpringerLink , and the ACM Portal were among the databases used. Additionally, Google Scholar was utilized to locate references for reviews of certain recommender system components. Search Descriptors: "Movie recommender systems," "movie personalization," "algorithms used in movie recommender systems," "filtering techniques in movie recommender systems," and "machine learning model metrics and measurement criteria" were a few of the keywords used to find information about movie recommender systems .Inclusion Requirements: Papers containing information regarding recommender systems required to come from published, peer-reviewed sources in order to meet the inclusion criteria. To ensure the accuracy of the data they included for use in this investigation, the publication abstracts were examined. Papers that contained grey literature on recommendation systems were excluded from consideration.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, May 2023

Machine learning algorithm for movie recommendation system

These are the algorithms that are employed in data mining and information filtering to provide the desired results. Understanding how information filtering techniques operate is crucial to choosing the appropriate algorithm for a given job in recommender systems.

How cosine similarity works

Cosine similarity is a statistic that is used to assess how similar two papers are, regardless of the size of the documents. The cosine of the angle made by two vectors projected onto a multidimensional space is computed. The cosine similarity is useful since it increases the likelihood that the two comparable documents will be orientated closer together, even if theyare separated by a large Euclidean distance because of the size of the documents. The cosine similarity increases with decreasing angle.

Cosine similarity is a mathematical concept that measures the cosine of the angle between two vectors in a multi-dimensional space. In the case of a movie recommendation system, each movie can be represented as a vector of attributes, and the cosine similarity between two movies is calculated as the cosine of the angle between their corresponding attribute vectors.

Here's an example of how cosine similarity could be used in a movie recommendation system:

The system builds a vector of attributes for each movie in the database, based on features such as genre, director, actors, and plot keywords.

When a user requests movie recommendations, the system takes the user's input and calculates the cosine similarity between that movie's attribute vector and the attribute vectors of all the other movies in the database.

The system returns a list of recommended movies, sorted by their cosine similarity scores.

Given two sentences:

A: I like watching TV, but I don’t like watching films.

B: I don’t like watching TV and films. How can we calculate the similarity between the two sentences? The basic idea is: the more similar the words used by the two sentences are, the more similar the sentences are.

First step: Word segmentation.

A: I / like / watching / TV, but / I / don’t / like / watching / films.

B: I / don’t / like / watching / TV / and / films. Second step: List all the words.

I, like, watching, TV, but, don’t, films, and Third step: Word frequency calculation. • A: I 2, like 2, watching 2, TV 1, but 1, don’t 1, films 1, and 0.

B: I 1, like 1, watching 1, TV 1, but 0, don’t 1, films 1, and 1.

Forth step: Get word frequency vector.

A: [2, 2, 2, 1, 1, 1, 1, 0]

B: [1, 1, 1, 1, 0, 1, 1, 1]

Then we can calculate the cosine of the two vectors by Equation 3.15.

$$\cos(\theta) = \frac{2 \times 1+2\times1+2\times1+1\times1+1\times0+1\times1+1\times1+0\times1}{\sqrt{2^2+2^2+2^2+1^2+1^2+1^2+1^2+0^2}\times\sqrt{1^2+1^2+1^2+1^2+0^2+1^2+1^2+1^2}}$$

The value is 0.85 so that the two sentences are much similar.

$$Cos\theta = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|\,\|\vec{b}\|} = \frac{\sum_1^n a_i b_i}{\sqrt{\sum_1^n a_i^2}\sqrt{\sum_1^n b_i^2}}$$

where, $\vec{a}\cdot\vec{b} = \sum_1^n a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$ is the dot product of the two vectors.

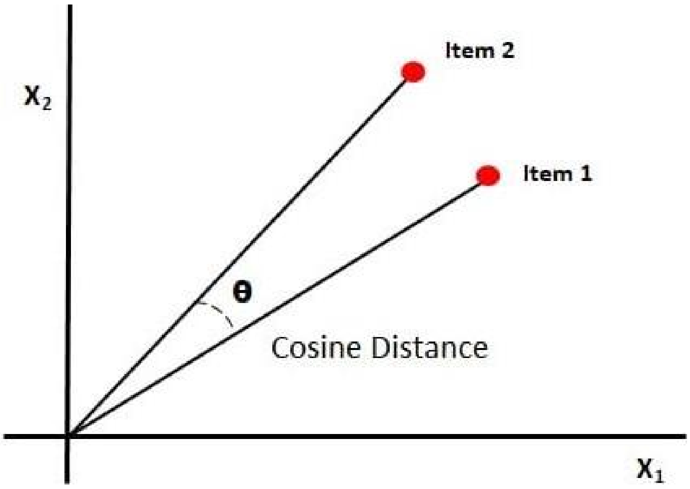

Fig : 2 work of Cosine Similarity

Watched by user

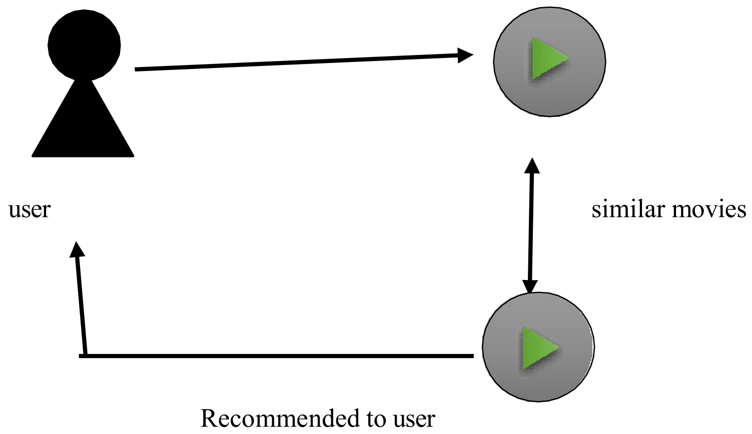

Fig : Content based filtering

Data collection The dataset for the project is the tmdb dataset which is available on kaggle

Movie Dataset :

- The data gathered from Kaggle has the following details:
- Title: Movie Title.
- Overview: Abstract of the Movie.
- Popularity: Movie popularity rating as per TMDB.
- Vote_average: Votes average out of 10.
- Vote_count: Number of votes from the users.
- Release_date: Date of release of the movie.
- Keywords: Keywords for the movie by TMDB in the list.
- Genres: Movie Genres in the list.
- Cast: Cast of the movie on the list.
- Crew: Crew of the movie in the list.
- Production_countries': It consists of the names of the countries in which the movie production took place.
- Release_date: It consists of the release date of the movie. The format used is yyyy-mm- dd where 'yyyy' indicates year of release, 'mm' indicates the month of release, and 'dd' indicates the day of release.
- Revenue: It indicates the revenue earned by the movie.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, May 2023

- Runtime: It indicates the runtime of a movie. Runtime basically means the length of the movie.
- Spoken_languages: It consists of the languages spoken in the movie.
- Title: It consists of the title of the movie.
- Vote_average: It indicates the average of the votes.

```
In [17]: movies.head()
Out[17]:
```

genres	homepage	id	keywords	original_language	original_title	overview	popularity	production_companies	productior
[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[{"name": "Ingenious Film Partners", "id": 289...	[{"iso_31("name": "U
[{"id": 12, "name": 'Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	139.082615	[{"name": "Walt Disney Pictures", "id": 2}, {"...	[{"iso_31("name": "U
[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...	en	Spectre	A cryptic message from Bond's past sends him o...	107.376788	[{"name": "Columbia Pictures", "id": 5}, {"nam...	[{"iso_31("nai
[{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,...	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112.312950	[{"name": "Legendary Pictures", "id": 923}, {"...	[{"iso_31("name": "U

Fig : Movie dataset

```
In [27]: movies = movies[['movie_id','title','overview','genres','keywords','cast','crew']]

In [28]: movies.head()
Out[28]:
```

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 1463, "name": "culture clash"}, {"id":...	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 470, "name": "spy"}, {"id": 818, "name...	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_id": "54805967c3a36829b5002c41", "de...
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	[{"id": 849, "name": "dc comics"}, {"id": 853,...	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
4	49529	John Carter	John Carter is a war-weary, former military ca...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 818, "name": "based on novel"}, {"id":...	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

Fig : movie main dataset column

```
In [38]: movies['genres'] = movies['genres'].apply(convert)
         movies['genres']

Out[38]: 0         [Action, Adventure, Fantasy, Science Fiction]
         1                          [Adventure, Fantasy, Action]
         2                            [Action, Adventure, Crime]
         3                       [Action, Crime, Drama, Thriller]
         4                   [Action, Adventure, Science Fiction]
                                          ...
         4804                           [Action, Crime, Thriller]
         4805                                   [Comedy, Romance]
         4806                    [Comedy, Drama, Romance, TV Movie]
         4807                                                  []
         4808                                       [Documentary]
         Name: genres, Length: 4806, dtype: object
```

Fig : Preprocessing the "genres " data in subsequent code.

```
In [164]: from sklearn.metrics.pairwise import cosine_similarity

In [165]: similarity = cosine_similarity(vector)

In [166]: similarity[0]

Out[166]: array([1.        , 0.15389675, 0.0860663 , ..., 0.        , 0.        ,
                 0.        ])
```

Fig : cosine similarity

```
In [173]: recommend('Avatar')

          Avatar
          Beowulf
          The Helix... Loaded
          Small Soldiers
          The Book of Life
```

Fig : Movie Recommendation

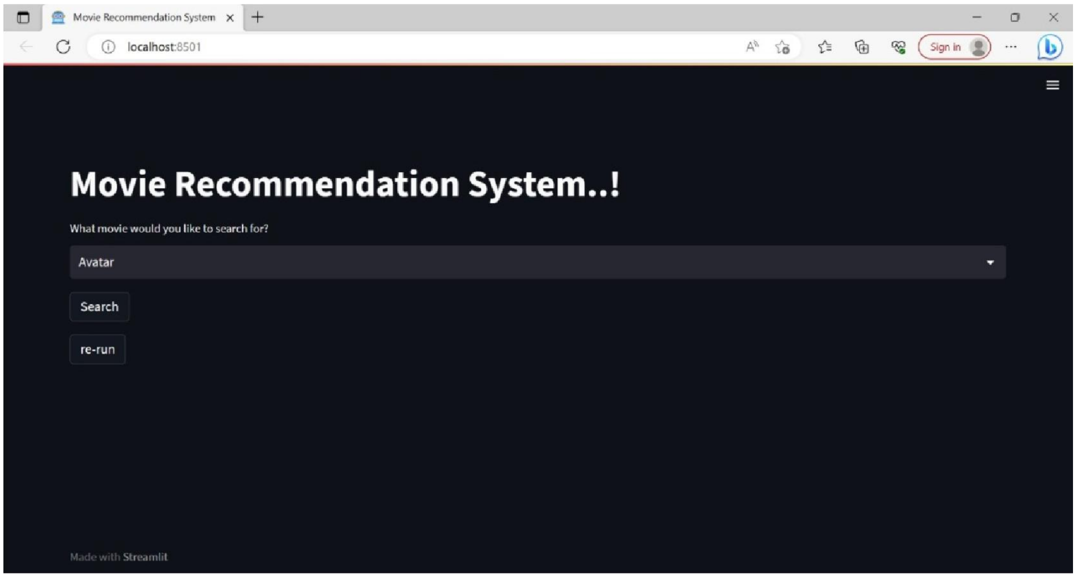

Fig : Output search movie

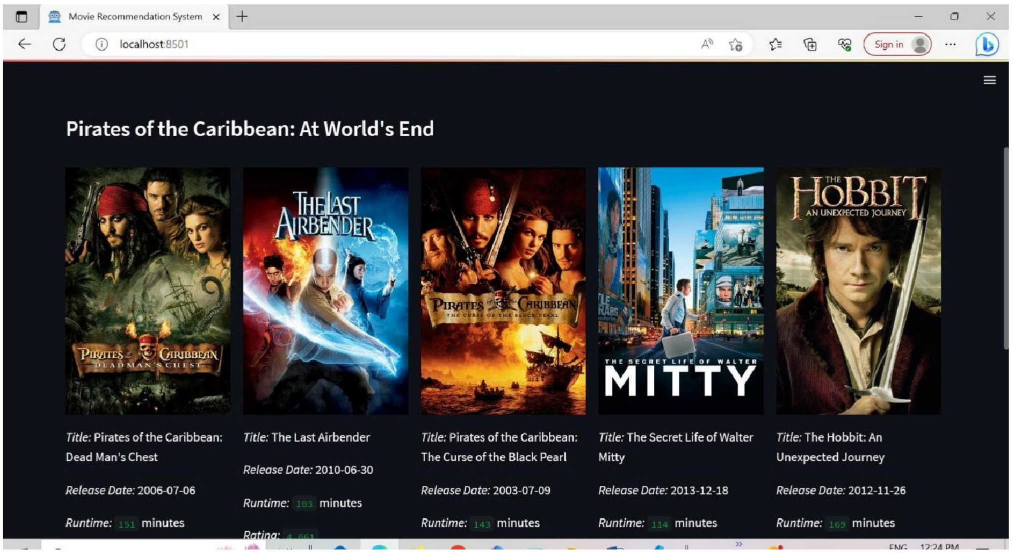

Fig : OUTPUT

III. FUTURE SCOPE

When we don't have enough user ratings for a movie or when user ratings for a particular movie are very high or low, the cosine similarity computation does not perform properly. Other approaches of computing similarity, including modified cosine similarity, can be utilized to improve this project.

The user vectors Ux and Uy are normalized before the adjusted cosine similarity, which is comparable to cosine similarity, is calculated by computing the cosine of the angle between them. However, unlike cosine similarity, adjusted cosine similarity substitutes the deviation between each user's raw item rating and their average item rating (denoted Ru) when computing the dot product of the two user vector.

IV. CONCLUSION

This essay is mainly split into two halves. One of them focuses on the sentiment analysis, while the other one is a movie recommendation system. The paper thoroughly examines both systems and draws some significant findings. The Cosine Similarity algorithm has been utilized for the Movie Recommendation System to suggest the best films that are relevant to the movie the user submitted based on several parameters such as the genre of the movie, overview, the cast, and the ratings provided to the movie. Even after multiple testing, Cosine Similarity has produced respectable findings and has been pretty reliable in terms of suggesting the films. In this study, sentiment analysis is also crucial. In essence, it seeks to categories the evaluations as favorable or unfavorable. For the same purpose, two algorithms have been utilized. The first is NB, while the second is SVC. Since there is a great deal of variation in reviews, it is crucial to select the optimal method for classification. This is the major motivation behind utilizing two algorithms to categories reviews. Finally, the experimental findings indicate that SVM has a very slight accuracy advantage over N. Here are a few outcomes of this study that have been mentioned:

1. Improving Sentiment Analysis's accuracy to better classify ironic or sarcastic evaluations.
2. Analysis of the reviews in other languages outside English's sentiment.
3. Personalized movie recommendations based on user preferences (cast, genre, year of release, etc.).

The technology has certain limitations even if it is quite precise. One of them is that the system won't suggest movies if the user-entered movie isn't included in the dataset or if the user doesn't input the name of the movie similarly to how it is in the dataset. The language barrier when performing the emotional analysis is still another drawback. Only reviews that have been written in English so far may be evaluated. If reviews are sardonic or caustic, the Sentimental Analysis incorrectly classifies them as well.

REFERENCES

[1]. Y. Rahhal, A. Jafar, and N. Nassar Deep multi-criteria collaborative filtering model for recommendations that is new Towards cognitive recommender systems: A. Beheshti, S. Yakhchi, S. Mousaeirad, S.M. Ghafari, S.R. Goluguri, and M.A. Edrisi

[2]. V. Rana, S. Sharma, and M. Malhotra Based on a hybrid filtering technique, an automatic recommendation system 1-16 in Educ. Inf. Technol., 27 (2021),

[3]. S. Chauhan, N. Rajput Various sentiment analysis methods are examined pp. 75–79 in Int. J. Computer Science and Mobile Computing, 8 (2) (2019).

[4]. A.A. Zulfiqar, M. Azeem, T. Mahmood, C. Xiao, Z. Shaukat, and Sentiment analysis on IMDB using neural networks and a lexicon, SN Appl. Sci., 2 (2) (2020), pp. 1–10.

[5]. K. De, S. Kumar, and P.P. Roy IEEE Trans. Comput. Soc. Syst., 7 "Movie recommendation system employing sentiment analysis from microblogging data" (4) A. Khan, M.E. Hossain,

[6]. S. Uddin, and M.A. Moni contrasting various illness prediction supervised machine learning methods

[7]. A. Adeel, K. Dashtipour, M. Gogate, H. Larijani, and A. Hussain Deep learning is used to analyse the sentiment of Persian movie reviews.

[8]. Soubraylu, S., and Rajalakshmi, R. Sentiment analysis using a hybrid convolutional bidirectional recurrent neural network for movie review.