



K. S. R. COLLEGE OF ENGINEERING, TIRUCHEGODE –
637 215
(Autonomous)

DEPARTMENT OF COMPUTER APPLICATIONS

**PROJECT:MOVIE RECOMMENDATION SYSTEM
USING SENTIMENT ANALYSIS FROM MICROBLOGGING
DATA**

Guided By:
Mr.N.PRAKASH
Assistant Professor

Register Number : 73152162040
Name : SANDEEP RAJ R

ABSTRACT

- Recommendation systems are important intelligent systems that play a vital role in providing selective information to users. Traditional approaches in recommendation systems include collaborative filtering and content-based filtering.
- However, these approaches have certain limitations like the necessity of prior user history and habits for performing the task of recommendation. In order to reduce the effect of such dependencies, of movie tweets.
- The movie tweets have been collected from microblogging websites to understand the current trends and user response of the movie.

EXISTING SYSTEM

- RS are mostly used in e-commerce applications and knowledge management systems such as tourism, entertainment and online shopping portals. In this paper, we focus on RS for movies are an important source of recreation and entertainment in our life.
- Movie suggestions for users depend on web-based portals. Movies can be easily differentiated through their genres like comedy, thriller, animation, and action.
- Another possible way to categorize movies can be achieved on the basis of metadata such as year, language, director or by cast.

DISADVANTAGE

- The primary requirement of a movie recommendation system is that, it should be very reliable and provide the user with the recommendation of movies which are similar to their preferences.
- In recent times, with exponential increase in amount of online data, RS are very beneficial for taking decisions in different activities of day-to-day life.

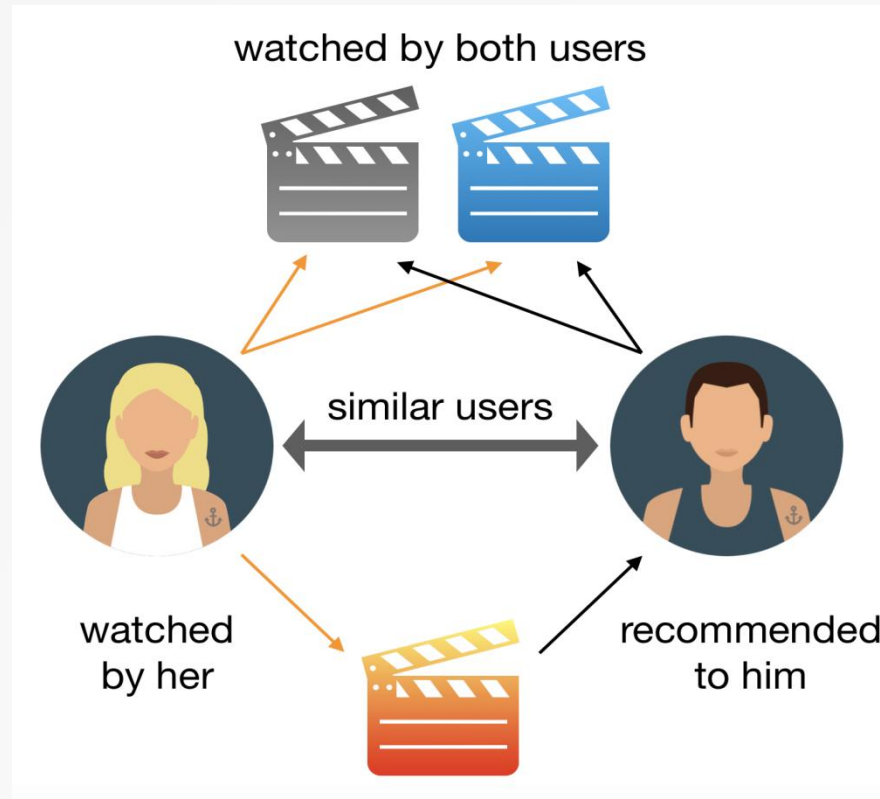
PROPOSED SYSTEM

- Propose a hybrid recommendation system by combining collaborative filtering and content-based filtering.
- Sentiment analysis is used to boost up this recommendation system.
- A detailed analysis of proposed recommendation system is presented through extensive experiment. Finally, a qualitative as well as quantitative comparison with other baselines models is also demonstrated

ADVANTAGE

- The Unique Selling Proposition of Twitter is that the existing users not only receive information according to their social links, but also gain access to other user-generated information.
- The source of information on Twitter are called tweets which are of limited-character that keep users updated about their favorite topics, people and movies.
- we propose a movie recommendation framework by fusing hybrid and sentiment scores from MovieTweatings database

SYSTEM ARCHITECTURE



SOFTWARE SPECIFICATION

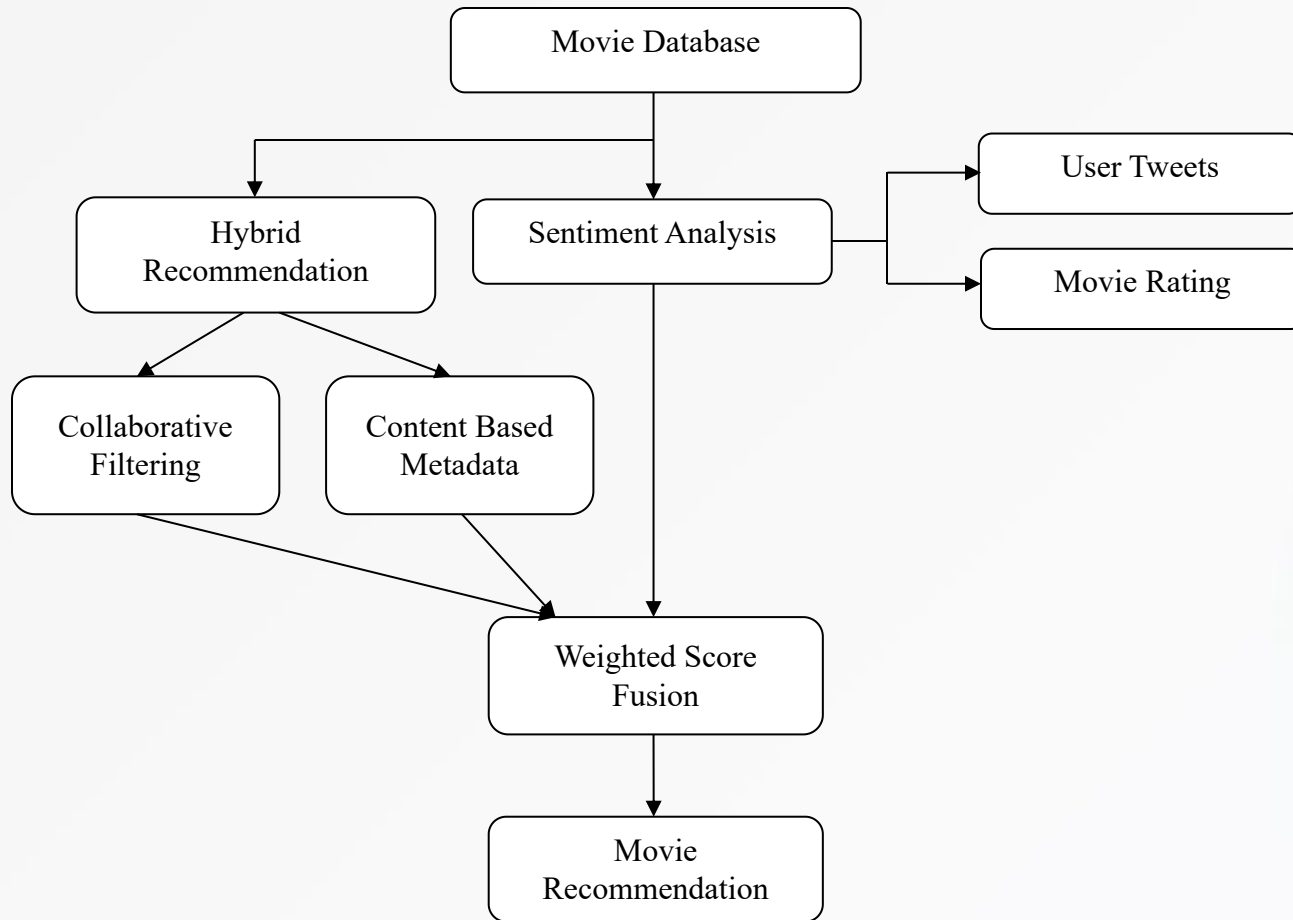
HARDWARE REQUIREMENTS:

- System : Pentium IV 2.4 GHz.
- Hard Disk : 500 GB.
- Monitor : 15 VGA Colour.
- Mouse : Logitech.
- RAM : 4 GB.


SOFTWARE REQUIREMENTS:

- Operating System : Windows-7/10 (64-bit).
- Language : Python 3.7
- IDE Tools : Visual Studio Code 1.7

DATAFLOW DIAGRAM



MODULES

- Data Set Collection
 - Analysis of User Tweets
 - Hybrid Recommendation
 - Weighted Score Fusion
 - Comparative Analysis Module
- 
- A decorative graphic on the right side of the slide, consisting of several overlapping, curved, wavy bands in shades of light blue, yellow, and a darker blue at the bottom right corner.

DATA SET COLLECTION

- **Public Databases:** There are many popular public databases available, which have been widely used to recommend the movies and other entertainment media. To incorporate the sentiment analysis in the proposed framework, the tweets of movies were extracted from Twitter against the movies that were available in the database.
- **Modified Movie Tweetings Database:** In the proposed work, the Movie Tweetings database is modified to implement the RS. The primary objective to modify the database was to use sentiment analysis of tweets by the users, in the prediction of the movie RS.

ANALYSIS OF USER TWEETS

- **Preprocessing of Tweets:** There are many short forms of words in the tweets, which converted into its original forms through gingerit8 library. To filter unusable data and uninformative parts in tweets such as stop words, punctuations, weblinks, and repetitive words, which did not add much value to sentiment analysis. After preprocessing, the text extracted from the tweets was used for sentiment analysis.
- **Sentiment Analysis of User Tweets:** VADER is a lexicon and rule-based method that is used to find the opinion expressed by the users in the form of tweets. It maps the words to sentiment by looking up the intensity of a word in the lexicon.

HYBRID RECOMMENDATION

- Describe the combination of content based similarity features with collaborative social filtering to generate a hybrid recommendation model. various hybrid recommendation techniques have been introduced and tested.
- Four major recommendation techniques constructing hybrids are collaborative filtering (CF), content-based (CN), demographic, and knowledge-based (KB). Unlike the first three which make use of learning algorithms, KB exploits domain knowledge and makes inferences about users' needs and preferences.

WEIGHTED SCORE FUSION

- To make the system robust, we use two data sources: one from the hybrid RS and another is from sentiment analysis. The hybrid RSs gives us the similarity between two movies based on their metadata (e.g., Actor, Director, Release Year, and Producer).
- The weights of these metadata for computing the similarity is computed under a linear system framework. The weights q signify the importance of particular metadata when a movie is compared with another (e.g., the genre of a movie has more importance than the runtime of the movie) movie.

COMPARATIVE ANALYSIS MODULE

- A comparative analysis of our proposed system with the pure hybrid model (PH Model) and sentiment similarity models (SS Models). The PH Model is a combination of CBF and CF.
- The recommended movies are based on the similarity of attributes, such as genre, director, and cast. This model is primarily effective when a fine-grained analysis is needed on the user's tweet along with its retweet to analyze multiple factors in a tweet, i.e., publisher, topic, and sentiment factors.

SCREENSHOT

```
In [3]: dataset.head()
```

```
Out[3]:
```

	Name	md5	Machine	SizeOfOptionalHeader	Characteristics	MajorLinkerVersion	MinorLinkerVersion	SizeOfCode	SizeOf
0	memtest.exe	631ea355665f28d4707448e442fbf5b8	332	224	258	9	0	361984	
1	ose.exe	9d10f99a6712e28f8acd5641e3a7ea6b	332	224	3330	9	0	130560	
2	setup.exe	4d92f518527353c0db88a70fddcfd390	332	224	3330	9	0	517120	
3	DW20.EXE	a41e524f8d45f0074fd07805ff0c9b12	332	224	258	9	0	585728	
4	dwtrig20.exe	c87e561258f2f8650cef999bf643a731	332	224	258	9	0	294912	

5 rows × 10 columns

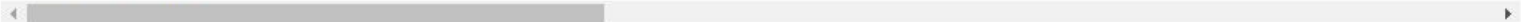
Recommendation Analysis

```
In [4]: dataset.describe()
```

```
Out[4]:
```

	Machine	SizeOfOptionalHeader	Characteristics	MajorLinkerVersion	MinorLinkerVersion	SizeOfCode	SizeOfInitializedData	SizeOfUninitializedData
count	138047.000000	138047.000000	138047.000000	138047.000000	138047.000000	1.380470e+05	1.380470e+05	1.380470e+05
mean	4259.069274	225.845632	4444.145994	8.619774	3.819286	2.425956e+05	4.504867e+05	1.009525e+05
std	10880.347245	5.121399	8186.782524	4.088757	11.862675	5.754485e+06	2.101599e+07	1.635288e+07
min	332.000000	224.000000	2.000000	0.000000	0.000000	0.000000e+00	0.000000e+00	0.000000e+00
25%	332.000000	224.000000	258.000000	8.000000	0.000000	3.020800e+04	2.457600e+04	0.000000e+00
50%	332.000000	224.000000	258.000000	9.000000	0.000000	1.136640e+05	2.631680e+05	0.000000e+00
75%	332.000000	224.000000	8226.000000	10.000000	0.000000	1.203200e+05	3.850240e+05	0.000000e+00
max	34404.000000	352.000000	49551.000000	255.000000	255.000000	1.818587e+09	4.294966e+09	4.294941e+09

8 rows x 55 columns



```
In [5]: dataset.groupby(dataset['legitimate']).size()
```

```
Out[5]: legitimate
0      96724
1      41323
dtype: int64
```

THANK YOU

A decorative graphic on the right side of the slide, consisting of several overlapping, curved, wavy bands in shades of light blue, yellow, and a darker blue at the bottom right corner.