A Project Report

on

Sentiments Based Movie Recommendation

Submitted in partial fulfillment of the

requirement for the award of the degree of

Bachelor of Technology in Computer Science and Engineering

Under The Supervision of
Mr. S.P.RAMESH
Assistant Professor
Department of Computer Science and Engineering

Submitted By

19SCSE1180050 - SAPAN MISHRA

19SCSE1180082 - NAITIK CHOUDHARY

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GALGOTIAS UNIVERSITY, GREATER NOIDA, INDIA

DECEMBER - 2021

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the project, entitled **" Sentiments Based Movie Recommendation "** in partial fulfillment of the requirements for the award of the **BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING** submitted in the **School of Computing Science and Engineering** of Galgotias University, Greater Noida, is an original work carried out during the period of **JULY-2021 to DECEMBER-2021**, under the supervision of **Mr. S.P. Ramesh, Assistant Professor**, **Department of Computer Science and Engineering** of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the project has not been submitted by me/us for the award of any other degree of this or any other places.

19SCSE1180050 - SAPAN MISHRA

19SCSE1180082 – NAITIK CHOUDHARY

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Supervisor

(Mr.S.P. RAMESH, Assistant Professor)

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of **19SCSE1180050 – SAPAN MISHRA, 19SCSE1180082 – NAITIK CHAOUDHARY** has been held on _____and his/her work is recommended for the award of **BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING**.

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date:

Place:

ABSTRACT

In the context of electronic commerce, recommender systems guide the user in a personalized way to interesting or useful objects in a large space of possible options. In order to provide reliable recommendation, the recommender systems need to exactly capture the customer needs and preferences into the user profile. However, for subjective and complexes products such as movies, music, news, user emotion plays surprising critical roles in the decision process. As the traditional model of user profile does not take into account the influence of user emotion, the recommender systems cannot understand and capture the constantly changing preferences of user. In this paper, we introduce an Emotion-based Movie Recommender System (E-MRS) as a solution to this problem. The objective of E- MRS is to provide adapted and personalized suggestions to users using a combination of collaborative filtering and content-based techniques. The recommendation is based on inferences about a user's emotions and preferences, as well as opinions of other similar users. This paper also discusses the system design and implementation, as well as its evaluation procedure. We believe that our system provides much better recommendation to users because it enables the users to understand the relation between their emotional states and the recommended movies.

Table of Contents

List of Figures

List of Tables

Acronyms

B.TECH	BACHELOR OF TECHNOLOGY
M.TECH	MASTER OF TECHNOLOGY
BCA	BACHELOR OF COMPUTER APPLICATION
MCA	MASTER OF COMPUTER APPLICATION
B.SC. (CS)	BACHLEOR OF SCIENCE
M.SC. (CS)	MASTER OF SCIENCE
SCSE	SCHOOL OF COMPUTER SCIENCE AND ENGIINEERING

CHAPTER-1

Introduction

ELECTRONIC Commerce is the movement of everything involving business on the Internet and the World Wide Web (www). Electronic Commerce will lead to simpler, faster and more efficient business transactions because the customers can benefit from the increasing range and ease of access to information, products and services.. The project is usefulfor saving valuable time and reduces the huge paper work. Recommender systems have become an answer to the need of personalization. The customer usually provides the recommender system with data such as the characteristics of the product he is looking for, his ratings, demographic data, etc. The recommender system applies one or several recommendation techniques on these data and then recommends products to the customers. In order to provide reliable recommendation, the recommender system needs to capture exactly the customer needs and preferences. However, for subective and complexes products such as movies, music, perfume, the task of rating or describing the desired product characteristics is quite difficult for customers. *M*oreover, as user preferences for these subjective products change constantly according to their emotions, the traditional user profile is not sufficient to understand and capture these changes. To solve these problems, we propose to use an Emotion-based Recommender System (E-MRS) that can capture customer preferences according to their emotions. Emotion plays an important role in rational and intelligent behaviour, thus, we incorporate user emotions in to the recommendation process.

1.1 Purpose

However, in today's competitive business environment, providing value to the customer is very important for businesses to survive. The most effective way to provide value is to know the customers and serve them as individuals. Customers need to feel they have a unique personal relationship with the business (Peppers and Rogers 1997]. Recommender systems have become an answer to the need of personalization. The customer usually provides the recommender system with data such as the characteristics of the product he is looking for, his ratings, demographic data, etc. The recommender system applies one or several recommendation techniques on these data and then recommends products to the customers. In order to provide reliable recommendation, the recommender system needs to capture exactly the customer needs and preferences.

1.2 Scope

For subective and complexes products such as movies, music, perfume, the task of rating or describing the desired product characteristics is quite difficult for customers. *M*oreover, as user preferences for these subjective products change constantly according to their emotions, the traditional user profile is not sufficient to understand and capture these changes. To solve these problems, we propose to use an Emotion-based Recommender System (E-MRS) that can capture customer preferences according to their emotions. Emotion plays an important role in rational and intelligent

behaviour, thus, we incorporate user emotions in to the recommendation.

1.3 Definitions :----

Personal details: Details of student such as user id, phone number, address, image, resume, e-mailaddress etc.

Electronic commerce---- movement of everything involving business

CHAPTER-2
LITERATURE SURVEY / PROJECT DESIGN

2.1 Literature Survey

The architecture of E-MRS (Fig. above.) is designed to recommend movies to users based on their emotions. To do so, the system captures user emotions by using a sequence of three colours. It will search for like-minded users who have similar emotion profiles and recommend movies with explanation. This system is a hybrid of two techniques:

collaborative filtering and content-based filtering.

2.2 Project Design

The purpose of the design phase is to develop a clear understanding of what the developer want people to gain from his/her project. As you the developer work on the project, the test for every design decision should be

"Does this feature fulfill the ultimate purpose of the project?"

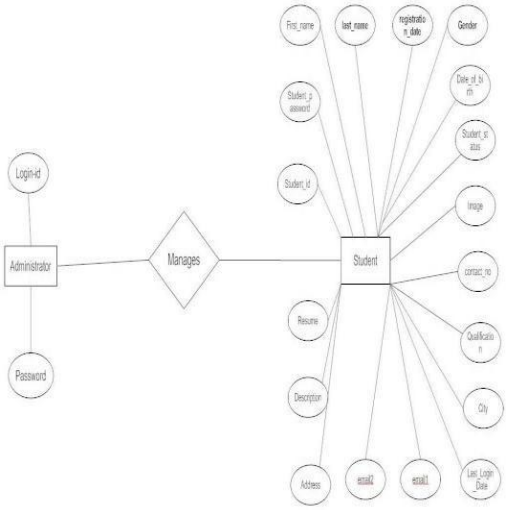
CORRELATION MEASURES BETWEEN SENTIMENT AND MOVIE RATINGS

Correlation coefficient	Definition	Value
PLCC	$\frac{\sum_{i=1}^{N}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{N}(y_i-\overline{y})^2}}$	0.76
SROCC	$1-\frac{6}{N(N^2-1)}\sum_{i=1}^{N} d_i^2$	0.72
KRCC	$\frac{2(N_c-N_d)}{N(N-1)}$	0.51

2.1 Entity Relationship Diagram

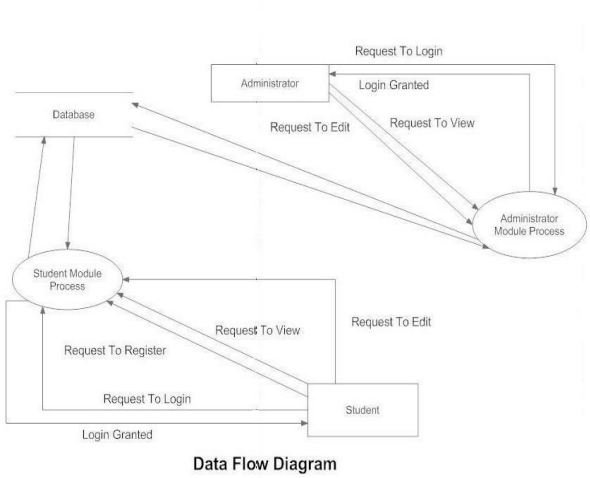
ER Diagram stands for Entity Relationship Diagram, also known as ERD is a diagram that displays the relationship of entity sets stored in a database. In other words, ER diagrams help to explain the logical structure of databases. ER diagrams are created based on three basic concepts: entities, attributes and relationships.

The SIMS is a system which contain major part which include: student Detail, Student image and resume.



2.2 Data Flow Diagram

Also known as DFD, Data flow diagrams are used to graphically represent the flow of data in a business information system. DFD describes the processes that are involved in a system to transfer data from the input to the file storage and reports generation



Data Flow Diagram

2.3 Decision Tree

A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g. whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The paths from root to leaf represent classification rules

CHAPTER 3

Functionality/Working of project

The proposed system needs two types of databases. One is a user-rated movie database, where ratings for relevant movies are present, and another is the user tweets from Twitter.

1) Public Databases: There are many popular public data-bases 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.

Experiments conducted using various public databases, such as the Movielens 100K,2 Movielens 20M,3 Internet Movie Database (IMDb,4) and Netflix database,5 that were not found suitable for our work due to the absence of microblogging data. After a thorough assessment of the abovementioned data-bases, the MovieTweetings database [12] was finally selected for the proposed system.

MovieTweetings is widely considered as a modern version of the MovieLens database. The purpose of this database

https://grouplens.org/datasets/movielens/100k/

https://grouplens.org/datasets/movielens/20m/

https://www.kaggle.com/orgesleka/imdbmovies

https://www.kaggle.com/netflix-inc/netflix-prize-data/data

TABLE I
DETAILS OF THE MOVIETWEETINGS DATABASE

Metric	Value
Ratings	646410
Unique Users	51081
Unique Movies	29228
Start Year	1894
End Year	2017

TABLE II
EXAMPLE OF A MOVIE ENTRY IN THE MODIFIED MOVIE TWEETINGS DATABASE

Attribute	Value
MovieID	0451279
Title	Wonder Woman
Runtime	141 min
Genre	Action,Adventure,Fantasy
Director	Patty Jenkins
Writer	Allan Heinberg
Actors	Gal Gadot',Chris Pine
Rating	7.6 Massachusetts Institute of Technology in 1996.
Production Companies	DC Films,Tencent Pictures
Popularity	524.772
Language	en
Production Countries	United States of America
Budget	816303142

is to provide an up-to-date movie rating so that it contains more realistic data for sentiment analysis. Table I displays the relevant details of the MovieTweetings database.

2) Modified MovieTweetings Database: In the proposed work, the MovieTweetings 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. The MovieTweetings database contains the movies with published years from 1894 to 2017. Due to the scarcity of tweets for old movies, we only considered the movies that were released in or after the year 2014 and extracted a

subset of the database which complied with our objective

release_yearmovies ≥ 2014.

The subset of the database consisted of 292 863 ratings by 51 081 users on 6209 different movies. The MovieTweetings database has three different components. The first component contains the mapping of users with their Twitter IDs. The second component contains the ratings of movies by users and their respective genres. The final component contains the information about the movies that were rated. In the proposed model, the socially filtered data, as well as the similarity of movies based on their attributes, has been used. The database had limited numbers of attributes for each movie, and thus, the Movie Database (TMDb) API was used to get more attributes of all the movies. TMDb6 is a premier source for extensive metadata for movies that have more than 30 languages. The movie attributes of the modified MovieTweetings database are showninTable II.

The modified database also contains some obscure movies from different countries and languages. The metadata for such movies was not available in TMDb, and therefore, those movies were discarded from the database. The final database had approximately 5000 movies.

TABLE III

EXAMPLES OF NOISY AND UNINFORMATIVE PARTS IN THE TWEETS

Types of noise	Example
Stop words	be, of, the, being, am
Stemmer	friend, friendship and friends
Web links	www.imdb.com
Filtering of repeating words	happyyyy, heloooo
Special Characters	!, @, #, $, %, and _

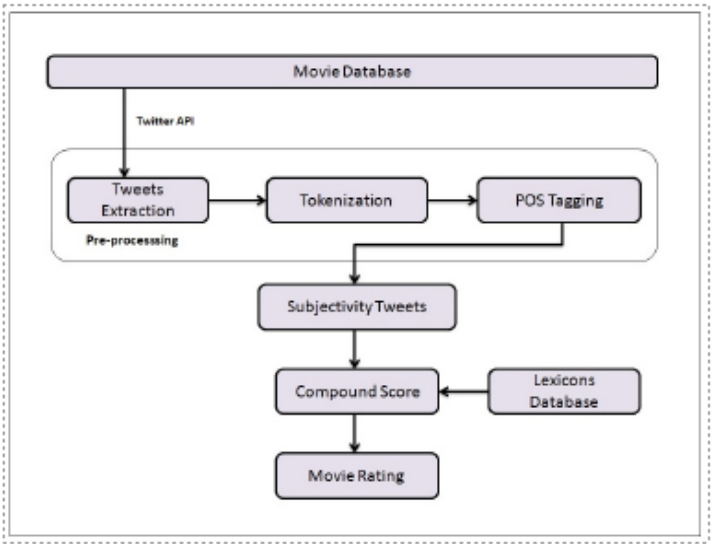


Fig. 2. Representative framework based on the VADER sentiment analysis system.

3.1 PROJECT CODING :

Python3 code for movie

recommendation based on

emotion

Import library for web

scrapping

```
from bs4 import BeautifulSoup as SOUP

import re

import requests as HTTP



# Main Function for scraping

def main(emotion):



        # IMDb Url for Drama genre of

        # movie against emotion Sad

        if(emotion == "Sad"):

        urlhere                                                                 =

'http://www.imdb.com/search/title?genres=drama&title_type=feature&sort=moviemeter, asc'



        # IMDb Url for Musical genre of

        # movie against emotion Disgust

        elif(emotion == "Disgust"):

        urlhere                                                                 =

'http://www.imdb.com/search/title?genres=musical&title_type=feature&sort=moviemeter,

asc'
```

```
        # IMDb Url for Family genre of
        # movie against emotion Anger
        elif(emotion == "Anger"):
        urlhere =
'http://www.imdb.com/search/title?genres=family&title_type=feature&sort=moviemeter, asc'

        # IMDb Url for Thriller genre of
        # movie against emotion Anticipation
        elif(emotion == "Anticipation"):
        urlhere =
'http://www.imdb.com/search/title?genres=thriller&title_type=feature&sort=moviemeter,  asc'

        # IMDb Url for Sport genre of
        # movie against emotion Fear
        elif(emotion == "Fear"):
        urlhere =
'http://www.imdb.com/search/title?genres=sport&title_type=feature&sort=moviemeter, asc'
```

```
        # IMDb Url for Thriller genre of
        # movie against emotion Enjoyment
        elif(emotion == "Enjoyment"):
        urlhere =
'http://www.imdb.com/search/title?genres=thriller&title_type=feature&sort=moviemeter,  asc'

        # IMDb Url for Western genre of
        # movie against emotion Trust
        elif(emotion == "Trust"):
        urlhere =
'http://www.imdb.com/search/title?genres=western&title_type=feature&sort=moviemeter,
asc'

        # IMDb Url for Film_noir genre of
        # movie against emotion Surprise
        elif(emotion == "Surprise"):
        urlhere =
'http://www.imdb.com/search/title?genres=film_noir&title_type=feature&sort=moviemeter,
asc'
```

```
        # HTTP request to get the data of

        # the whole page

        response = HTTP.get(urlhere)

        data = response.text



        # Parsing the data using

        # BeautifulSoup

        soup = SOUP(data, "lxml")



        # Extract movie titles from the

        # data using regex

        title = soup.find_all("a", attrs = {"href" : re.compile(r'\/title\/tt+\d*\/')})

        return title



# Driver Function

if__name__== '_main__':
```

```
emotion = input("Enter the emotion: ")

a = main(emotion)

count = 0



if(emotion == "Disgust" or emotion == "Anger"

                                or emotion=="Surprise"):



for i in a:


        # Splitting each line of the

        # IMDb data to scrape movies

        tmp = str(i).split('>;')



        if(len(tmp) == 3):

                print(tmp[1][:-3])



        if(count > 13):

                break
```

```
        count += 1
else:
for i in a:
        tmp = str(i).split('>')


        if(len(tmp) == 3):
                print(tmp[1][:-3])


        if(count > 11):
                break
        count+=1
```

CHAPTER 4

RESULT

We conducted the statistical analysis between sentiment ratings X and movie rating Y to find the correlation coefficient. The correlation coefficient value varies from −1to +1. Let D denotes a database of movies and N denote the number of total movies in the database. The statistical correlation coefficients are as follows: Spearman rank-order correlation coefficient (SROCC), Kendall rank correlation coefficient (KRCC), and Pearson linear correlation coefficient (PLCC). Table IV dis-plays the values of different corelation coefficients utilized by us. In our experiments, we have found that sentiment and movie ratings are positively correlated. For PLCC, xi and yi are sentiment rating and IMDb movie rating, respectively, for the ith movie, whereas $\bar{x}$ denotes the mean sentiment score and $\bar{y}$ denotes the mean movie rating in the database. For SROCC, di is the difference between the sentiment rating and movie rating of the ith movie in the database. For KRCC, Nc and Nd represent the number of concordant and discordant pairs in the database, respectively.

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

5.1 CONCLUSION

RSs are an important medium of information filtering systems in the modern age, where the enormous amount of data is readily available. In this article, we have proposed a movie RS that uses sentiment analysis data from Twitter, along with movie metadata and a social graph to recommend movies. Sentiment analysis provides information about how the audience is respond to a particular movie and how this information is observed to be useful. The proposed system used weighted score fusion to improve the recommendations. Based on our experiments, the average precision in Top-5 and Top-10 for sentiment similarity, hybrid, and proposed model are 0.54 and 1.04, 1.86 and 3.31, and 2.54 and 4.97, respectively.

We found that the proposed model recommends more precisely than the other models.

5.2 FUTURE SCOPE

In the future, we plan to consider more information about the emotional tone of the user from different social media platforms and non-English languages to further improve the RS.

References

[1] Burke, R. “Knowledge-based Recommender Systems” . In: A. Kent (ed.): Encyclopedia of Library and Information Systems. Vol 69. 2000, pp. 2-23.

[2] Schafer, J. B. Konstan, J. A. Riedl, J. “E-Commerce Recommendation Applications”, In Data Mining and Knowledge Discovery. Vol 5. 2001, pp. 7-39

[3] P. Resnick, N. Iakovou, M. Sushak, P. Bergstrom, and J. Riedl, ” GroupLens: An Open Architecture for Collaborative Filtering of Netnews," Proc. 1994 Computer Supported Cooperative Work Conf., 1994.

[4] Matthew Garden, Gregory Dudek. “Mixed Collaborative and Content-Based Filtering with User-Contributed Semantic Features" . In, AAAI National Conference on Artificial Intelligence, Boston, Massachusetts. 2006, pp. 1-6

[5] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey "Toward machine emotional intelligence: analysis of affective physiological state" .In, EEE Transactions on Pattern Analysis and Machine Intelligence. Vol 23 .2001, pp. 2-4

[6] Hill, W., Stead, L., Rosenstein, M., And Furnas, G. W “Recommending and evaluating choices in a virtual community of use". In Proceedings of ACM CHI 95 Conference on Human Factors in Computing Systems. 1995, pp 194-201.

[7] Goleman. Emotional intelligence. New York : Bantam Books, 1995

PUBLICATION

ACCEPTED IN ARDA PUBLICATION

TURCOMAT

Turkish Journal of Computer and Mathematics Education

E-ISSN 1309-4653

Email : info@ardaconference.com

Website : www.ardaconference.com/journal.php

Mobile : +91 93456 84472 / Whatsapp : +91 93456 84472

Date: 5th Dec 2021

Acceptance Letter

Dear Author(s): SAPAN MISHRA, NAITIK CHAUDHARY

Paper ID	ARDA_Publication_8394
Paper Title	Sentiments Based Movie Recommendation

This is to enlighten you that above manuscript reviewed and appraised by the review committee members and it is accepted for the purpose of publication in the **"Turkish Journal of Computer and Mathematics Education"** with **ISSN: 1309-4653** that will be available at https://dergipark.org.tr/en/pub/turkbilmat

You have to send following documents at info@ardaconference.com on or before 19th Dec 2021.
1. Final Paper | Ms Word .doc/.docx file.
2. Proof of Payment | Scanned | Online Received Email.

Note: Please read carefully

1. Above manuscript will be published within 30-45 days from your payment.
2. **Turkish Journal of Computer and Mathematics Education** is a Scopus Indexed Journal.
3. Author(s) will receive Publication information and Published Paper link through ARDA.
4. You may see more about the journal at: https://dergipark.org.tr/en/pub/turkbilmat
5. You will receive Volume/ Issue information of your paper very soon.

Sincerely

Dr. Simpson Rodricks
President,
ARDA.