

# Movie Recommendation System

Supreeth Suryaprakash, Suhas Jagadish, Raghuveer Krishnamurthy and Srikanth Srinivas

Dept. of Computer Science  
Indiana University, Bloomington, United States  
skeragod@indiana.edu  
jagadiss@iu.edu  
rkanchib@iu.edu  
sriksrin@iu.edu

**Abstract—** In our system we recommend movies to the user based on their YouTube comments. We accumulate each user's YouTube comments and we make use of the sentiment classifier to generate the positive and negative score for that comment. Based on the positivity of his comments we determine his preferred genre. Our system recommends YouTube movies/videos related to same genre, top 10 IMDB movies related to the same genre and twitter trends related to the recommended movies from YouTube and IMDB.

approach uses the preferences of other people with similar tastes for recommending items to the user.

We have developed a movie recommendation system based on the user's comments on YouTube videos and movies. This system uses a hybrid approach in recommending movies in a way that it uses user's information (content based) from his/her comments, recommends similar (collaborative) movies/videos in same genre.

## I. INTRODUCTION

With recent advancements in web and its connectivity with millions of people across the whole world, enormous amount of information is generated daily which are easily accessible to the users. In the flood of such information, users want to have means to search for desired information easily and quickly. To help the users cope with this information explosion, various recommendation systems have been developed to guide the users for movies such as Jinni, Rotten Tomatoes, IMDB and Netflix. However, these systems do not seem to offer help with the user's decision-making due to their inability to reflect the characteristics of individual users. Personalized services mean providing consumers with services most suitable to individual tastes based on their personal information, activities and their opinions. Although aforementioned systems provide services based on user's tastes and activities, user's opinions amalgamated in their comments have not been taken into consideration. Our system primarily focusses on user's comments.

Recommendation systems are special types of expert systems in the sense that they combine the knowledge of the expert in a given domain (for the service being recommended) with the user's preferences to filter the available information and provide the user with the most relevant information. Personalization of the recommendations works by filtering a candidate set of items (such as web pages) through some representation of a personal profile. Two main paradigms for the filtering are content-based approach and collaborative approach. Most recommendation systems use a hybrid approach, which is a combination of these two approaches. A content based recommendation system uses the user's past history to recommend new items where as a collaborative

## II. BACKGROUND AND RELATED WORK

Over the past decade, a large number of recommendation systems for a variety of domains have been developed and are in use. Most of the online recommendation systems for a variety of items use ratings from previous users to make recommendations to current users with similar interests. One such system was designed by Jung, Harris, Webster and Herlocker (2004) for improving search results. The system encourages users to enter longer and more informative search queries, and collects ratings from users as to whether search results meet their information need or not. These ratings are then used to make recommendations to later users with similar needs.

Quijano-Sanchezat.al, 2011 have been issued a movie recommender system as an application in name Happy Movie (HM), for group of people which it is integrated with the social network face book (FB). They tried through HM to diminish a certain limitations in existing group RSs, like obtaining the user's profile or offering trading methods for users in order to reach a final agreement. The utilized method to make the group recommendation is based on three important features: personality, social trust and memory of past recommendations. Eventually, they simulate in a more realistic way the argumentation process followed by groups of people when deciding a joint activity.

Menezes, and Tagmouti, introduced 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 CF 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. In their paper discuss 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 in their comments and the recommended movies.

MovieLens is an online movie recommendation system. When a user logs in for the first time, the user is asked to rate certain movies that the user has seen. These ratings are then used to recommend other movies to the user that the user has not seen. It also uses collaborative filtering based on ratings by similar. These two approaches are combined to create personalized recommendations.

Rajenderan, A, 2014 complete a project is to create a movie RS that uses human emotion as the basis of recommendation. The system will observe a user while they watch a portion of content, and then analyze the data that is their facial expressions and heart rate over time. With a sufficient database, their emotion data can be compared with those of other users, and recommendations can be provided based on reactions to the content [13], furthermore explains the landscape of actual and possible hybrid recommenders, and introduces a novel hybrid, a system that combines content boosted recommendation and CF to recommend restaurants.

### III. MOTIVATION

There are 2 main reasons which motivated us to develop this system. First, User gets narrowed down results from multiple domains on one web page. The user might provide a positive comment on a YouTube video, but might like to know the reviews of movies in the same genre from IMDB and also would like to know the opinions of other users in twitter. Current recommendation systems analyze user's sentiments based on his/her activities in one website and recommend related videos/movies in the same website. For E.g. YouTube recommends videos in its own database based on user's activities.

Second and the most important, YouTube recommends videos/movies based on users' recent browsing history and not on the user's sentiment in comments. Even if the user has provided negative comment on certain videos in one genre, the user is recommended with the videos of same genre. Our system addresses this gap by analyzing the user's sentiment to recommend him/her with similar videos.

### IV. PROPOSED METHODOLOGY

Crawl YouTube to gather user's comments and store in Database. Analyze user's sentiment using sentiment classifier, which provides overall positive and negative score of each comment, based on these score, rating will be calculated. For top rated videos/movies, genres are selected. IMDB Dataset is used, stored in database. Based on top rated movies/videos, the same genre similar IMDB movies are recommended for the user. Also, Twitter trends on all recommended movies would be shown by using Twitter APIs.

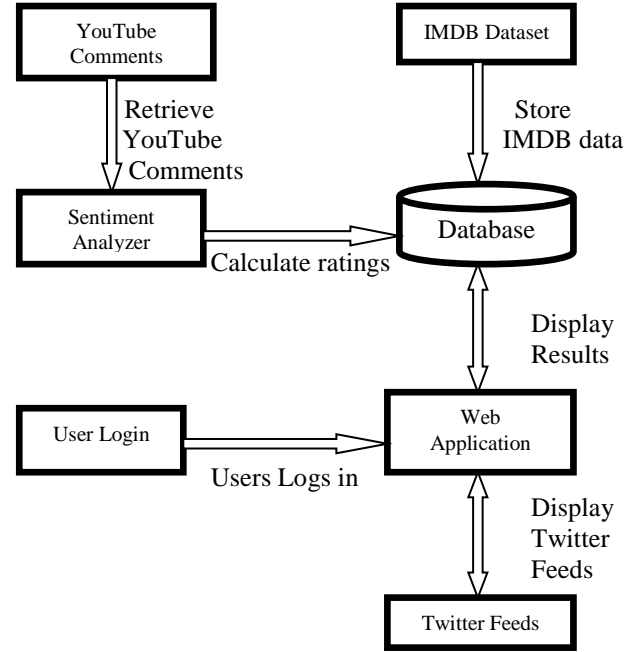


Fig 1: System Design

### V. IMPLEMENTATION

Our data set comprises of comments mined from YouTube for specific set of users and videos using YouTube Data API. We also collected data related to IMDB movies from IMDB data set. Collected data is modelled into 4 tables - User, Video, Movie and Rating using MySQL. The User table comprises of user information such as userID, username, videoID and comments. Video table contains videoID and genre information. Movie table contains IMDB movies, genre and rating information. The system then performs sentiment analysis on YouTube comments and predicts a rating for each comment using our rule-based classifier. The Rating table is then updated with ratings for each comment.

Using the senti wordnet files, the system then first assigns weights to each word in the wordnet based on the intensity. For example, a positive word "Awesome" would have a weight of 10 whereas another positive word "good" will have a weight of 6.

The rule-based classifier looks at the weights associated with each word and the number of positive or negative words in a comment and assigns a positive and negative score for each comment. Based on the percentage of positive score, each comment is associated with a rating on a scale of 0-5, with 5 being highly positive and 0 being highly negative comment.

We performed 5-fold cross validation to measure the accuracy of our rule-based classifier. We picked 1000 comments from YouTube and manually labelled ratings for each comment. This will be the golden standard for our analysis. Using 5-fold cross validation we compared the accuracy of our rule-based classifier with the python in-built senti classifier.

The results are shown in the below table –

	Our classifier	Senti classifier
Set 1	70.67	80.4
Set 2	72.13	82.3
Set 3	68.27	80.1
Set 4	71.45	82.3
Set 5	72.06	83.8

We achieved an overall accuracy of 70.92% as compared to 81.67% accuracy of python senti classifier. This is a significant accomplishment since our classifier is primitive and does not detect sarcastic comments as compared to the python senti classifier. This is one of the components of our future enhancements - to detect sarcasm in comments.

For each user we pick the genres where users' rating is the highest and recommend the YouTube videos and top rated IMDB movies of the same genre. We have built an interactive web application using PHP which is used for visualization of the results.

When the user logs on to the application, he will be first authenticated. For the user login authentication, the system uses an MD-5 encryption technique. The password will be decrypted in a function written in the LOGIN php script. Once, the User logs in, we are storing the user details in a session object which will be destroyed only when he logs off. Typically, if the user opens multiple tabs in a single browser, he will be authenticated into the system automatically. If the user, opens the same portal from a different browser, this session will be destroyed and the user will be asked to login again in this browser. The system will store only one session per user at any point of time. When the login is successful, user will be presented with a list of YouTube videos and IMDB movies of his favorite genre. User can watch the YouTube video on the go or click on the IMDB movie link which redirects the user to the corresponding IMBD page.

Twitter feeds have been integrated with the system. It will display the tweets/trends about the recommended movies. Below are the few snapshots of the application



Fig 2: Login page



Fig 3: Recommended YouTube Movies and Twitter feeds



Fig 4: Recommended IMDB Movies and Twitter feeds

## VI. CONCLUSION AND FUTURE ENHANCEMENTS

In this paper we designed a new web recommender system for movies based on user’s sentiment in their YouTube comments. Rule-Based classifier on YouTube comment’s on training data constitute the core part of our system. The system uses rule based classifier for calculating the overall positive and negative score of each comment and calculates the rating. Finally the users are suggested YouTube videos, IMDB movies in their preferred genre and trending tweets from Twitter in an interactive web application. A larger data set will enable more meaningful results using this system. Additionally we would like to incorporate different machine learning algorithms and study the comparative results. We can extend the current system to integrate other online movie systems such as Netflix and Hulu. For large datasets we would like to use Hadoop and MapReduce to move the datasets to the Cloud.

## REFERENCES

- [1] Ricci. F, Rokach. L and Shapira. B, "Introduction to recommender systems handbook": Springer
- [2] G. Eyrun A. Eyjolfssdottir, Gaurangi Tilak and Nan Li, "MovieGEN: A Movie Recommendation System"
- [3] Karzan Wakil, Rebwar Bakhtyar, Karwan Ali and Kozhin Alaadin, "Improving Web Movie Recommender System Based on Emotions"
- [4] Sajal Halder, A. M. Jehad Sarkar and Young-Koo Lee, "Movie Recommendation System Based on Movie Swarm"

- [5] Michael Fleischman and Eduard Hovy, "Recommendations Without User Preferences: A Natural Language Processing Approach"
- [6] Lawrence R.D, Almasi G.S, Kotlyar V, Viveros M.S and Duri S, "Personalization of supermarket product recommendations(2001)"
- [7] Quijano-Sanchez. L, Recio-Garcia. J. A and Diaz-Agudo. B, "Happymovie: A facebook application for recommending movies to groups"
- [8] Rajenderan. A, "An Affective Movie Recommendation System"
- [9] M. Balabanovic, "An adaptive web page recommendation service"
- [10] Sajal Halder, Md. Samiullah A. M, Jehad Sarkar and Young-Koo Lee, "Movieswarm: Information mining technique for movie recommendation system"
- [11] Christakou. C, Lefakis. L, Vrettos. S and Stafylopatis. A, "A Movie Recommender System Based on Semi-supervised Clustering"
- [12] Qing Li and Byeong Man Kim, "Clustering Approach for Hybrid Recommender System"
- [13] J. B. Schafer, J. Konstan and J. Riedl, "Recommender systems in e-commerce"
- [14] Deuk Hee Park, Hyea Kyeong Kim, Il Young Choi and Jae Kyeong Kim, "A Review and Classification of Recommender Systems Research"
- [15] Basu. C, Hirsh. H, and Cohen. W, "Recommendation as classification: using social and content-based information in recommendation".

## APPENDIX

### Suggested enhancements by professor

- 1) Inclusion of actions in architecture design diagram
- 2) Integration of Netflix with current system
- 3) Suggest users to follow with similar genre interests
- 4) Integration of twitter feeds into the system, which has been completed

### Classes Missed

Supreeth – 0

Raghuveer – 0

Srikanth – 1[Arrived late, informed professor]

Suhas – 1[Arrived late, informed professor]