

Project Proposal

IMDB Rating Prediction Based On YouTube Trailer Data

Adarsh Rahul Jaju (173050003)

Amit Patil (173050009)

Kalpesh Dusane (173050042)

Dhaval Limdiwala (173050061)

CS-725

IIT Bombay

September 20 2017

Contents

1	Project description	3
2	Approaches	3
3	Research papers	4
4	Data sets	4

1 Project description

Motivation : The current IMDb rating is calculated after the realease of a movie, so viewers have to wait for the movie to be released to check whether the movie is good by cheking the IMDb rating of the movie. What if one has to book tickets prior to the relaese of a movie. So this motivated us to calculate movie ratings based on the movie trailer relaesed on Youtube.

We are going to predict IMDb rating based on data of trailer video of a movie which is on YouTube like number of views, number of likes, dislikes, movie genres(like Adventure, Thriller, Horror, romantic,comedy), time of the year, is it franchise based?(like Fast & Furious, Marvel cinematic universe), movie cast(like Director, Actors and actress). Also we are Analyzing YouTube Comments. The main idea is to use linear regression model.

2 Approaches

1. Linear Regression : To construct the features based on Youtube trailer data(different features mentioned above). With the help of this data, model will be created. Then using linear regression we will find the weight vector for our model.
2. Neural Network : Till now we have some rough idea about Neural Networks, if it fits our problem statement better than linear regression, we will be using it.
3. Sentiment analysis : We are planning to use machine learning algorithms to handle sarcastic sentences and use nltk library for stemming of the words(remove unnecessary/meaningless words like a,the etc. And to classify different forms of a word to the basic form of the same word, like love, lover, loving will match to same word form love)
4. Part of speech tagging : We are planning to use Stanford Part of Speech Tagger, which has 98.5 % accuracy for their left three words model, with which we can analyze nouns, adjectives of a sentence, which may help in the sentiment analysis while analyzing the comments.

3 Research papers

1. Text Analysis of YouTube Comments
<https://www.curiousgnu.com/youtube-comments-text-analysis>
2. Predicting IMDB Movie Ratings Using Social Media
https://www.researchgate.net/publication/262401833_Predicting_IMDB_Movie_Ratings_Using_Social_Media
3. Adaptive sentence boundary disambiguation
<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.14.1553&rep=rep1&type=pdf>

4 Data sets

We need the features like number of views, number of likes, number of dislikes, Movie Genres, movie cast, comments(classified as positive, negative and neutral) in data set.

1. <http://netsg.cs.sfu.ca/youtubedata/>
2. We are planning to extract data from youtube using web scrapping.