

Intelligent Prediction of Ratings & Profit Estimation for Upcoming Movies

SNEHAL PATEL: A20329879

ADITYA KULKARNI: A20332813

Problem Statement

- Success of a movie depends on multiple factors such as cast, director, producer or genre but there is lack of analytics for predicting movie's success in terms of rating and revenue.
- Producers and Investors find it difficult to estimate the revenue based on current expenses involved in production.
- Influence of Twitter and YouTube on the rating of an upcoming movie.



People cannot decide if going for an upcoming movie is worth spending the money !!

Approach for Twitter & YouTube analysis:



- Used sentiment140 API for Sentiment analysis of trending tweets of an upcoming movie to generate positive and negative scores. Further, we fit a logistic regression classifier to predict the rating of the upcoming movie.
- We analyzed the likes and dislikes of the official YouTube trailer of the upcoming movie and generated positive and negative scores.

Approach for prediction model:

Initial input features	
Features	Weights
Actors	3.5
Directors	3
Genre	2
Production House	1.5
Total	10

- Built a prediction model based on weighted quantitative measurement to predict the rating of upcoming movie.
- Estimating revenue generated by upcoming movie.
- Calculating the score/ratings of movie based on category's weight.

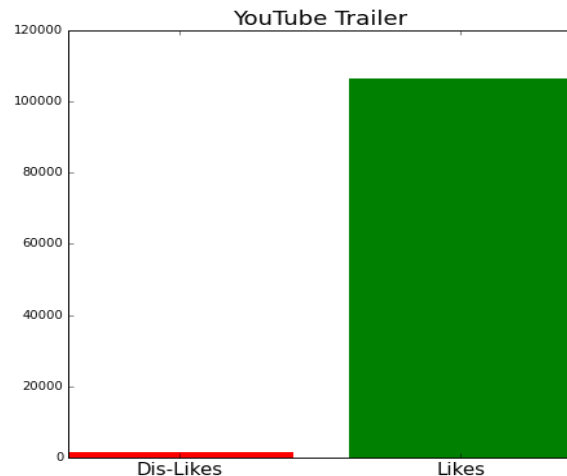
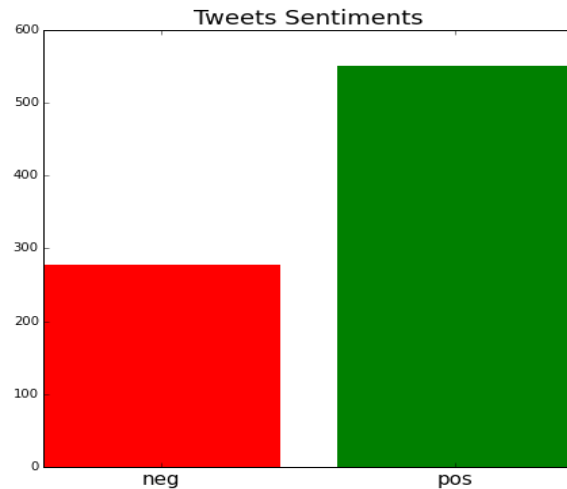
Data Collection

Data sources:

- IMDb API:
 - Data collection of approximately 8k movies from year 2000 to 2015.
 - Each movie has 34 different parameters of information.
 - We used the following movie metrics:
 - Star cast , Director, Producer, Genre, Ratings from IMDb, Rotten Tomatoes ratings.
- YouTube API:
 - Data collection of upcoming movie's official trailer.
 - Data included, view counts, likes, dislikes.
- Twitter API:
 - Data collection of 10k tweets for any upcoming movie.



Result of Twitter & YouTube analysis



- Used Sentiment140 API to classify upcoming movie's tweets as positive, negative, and neutral tweets.
- The accuracy obtained by applying Logistic Regression using 7 folds cross validation is about 65%.
- We have extracted likes & dislikes from upcoming movie's official trailer as a prediction metric.

Result of prediction model

$$Score = \sum_{i=1}^n M_i * W_i / \sum_{i=1}^n W_i$$

m_i – score of a movie that artist was in
 W_i – contribution of artist p in movie m_i

- At first we calculated rating based on initial weights on released movie which resulted in a low accuracy.
- Initial weights = [3.5, 3, 2, 1.5] Accuracy = 63.22%
- To improve the accuracy we optimized the initial weights
- Optimized weights = [3.8, 2.7, 3.1, 0.4] Accuracy = 82.53%

Conclusion

- Based on a comparative analysis of accuracies of individual ratings from three different approaches and the accuracy of combined rating of all the approaches, we can conclude that Twitter & YouTube influences the rating of any upcoming movie.
- As future work, we can use multiple better classifiers to further improve the accuracy.