

MACHINE LEARNING TERM PROJECT PROPOSAL

Project Title:

Predicting Box Office Revenue using Social Networks and Trends

Project Idea:

Given the social media buzz the movie is generating and the people involved in the movie production, can we predict the revenue it'll generate once it's out?

The idea of our project is to predict the box office collections of an upcoming movie using the attributes like the actors, director and producers of the movie (from IMDB), how popular the actors appearing in the movie are (using their social media following) from Twitter and Facebook, budget of the movie (from Wikipedia), number of views and comments on the trailer on YouTube and the buzz the movie is generating based upon the searches related to it using data from Google Trends.

Data Set:

Data would be procured from various sources namely - Wikipedia, IMDB, YouTube, Twitter, Facebook and Google Trends.

Software:

Python, Scikit-learn libraries of python, Weka, Ruby on Rails for the Web interface

Project Timeline:

30th March - Data collection and cleaning after API integration
10th April - Implementation of algorithms
18th April - (optional) Web based UI implementation
20th April - Draft project report
30th April - Submission

References:

- Using Sentiment and Social Network Analyses to Predict Opening-Movie Box-Office Success by Lyric Doshi http://web.mit.edu/lyric/Public/Lyric_Doshi_M-Eng_Thesis_Feb_2010.pdf
- Early Predictions of Movie Success: the Who, What, and When of Profitability Michael T. Lash and Kang Zhao <https://arxiv.org/pdf/1506.05382v2.pdf>
- Improving Movie Gross Prediction Through News Analysis, Wenbin Zhang and Steven Skiena Department of Computer Science Stony Brook University. <http://www3.cs.stonybrook.edu/~skiena/lydia/movie.pdf>

Teammates

1. Ashish Mohapatra (axm160031)
2. Dhruv Gajanan Sangvikar (dgs160230)
3. Vipul Nataraj Jarmale (vxj160130)