Multiple Regression Case Study Description

Chen Nan

September 21, 2017

1 Predict Movie Rating

How can we tell the greatness of a movie before it is released in cinema? There seems to no universal way to claim the goodness of a new movie. Many people rely on critics to gauge its quality, while others use their instincts after watching trailers or seeing the poster in the subway. However, it takes the time to obtain a reasonable amount of critics reviews after a movie is released, and human instinct sometimes is unreliable. Is there a better way for us to tell the greatness of movies without relying on critics or our own instincts?

To answer this question, someone scraped 5000+ movies from IMDB website using a Python library called "scrapy". In the end, I was able to obtain all needed 28 variables for 5043 movies and 4906 posters, spanning across 100 years in 66 countries. There are 2399 unique director names, and 30K actors/actresses.

The image below shows all the 28 scraped variables. Roughly speaking, half of the variables is directly related to metadata of movies themselves, such as title, year, duration, etc. Another half is related to the people who involved in the production of the movies, eg, director names, director facebook popularity, movie rating from critics, etc.

How does the IMDB rating score correlate with other variables? Gross, actors, directors, genr, etc? Movies with rating score higher than 8.0 are listed in the IMDB top 250, and they are considered as truly great movies by critics from many perspective. Movies with rating score from 7.0 to 8.0 are probably still good movies. Viewers can gain something from them. On the contrary, movies with rating scores from 1 to 5 are sometimes considered as ones that "sucks", in one way or the other. One perhaps should avoid those movies unless they have to.

What are the predictors that are relevant to make forecasting for this purpose (before it is released)? Can we make predictions on the box revenues based on these information?

See attachment: movie.csv

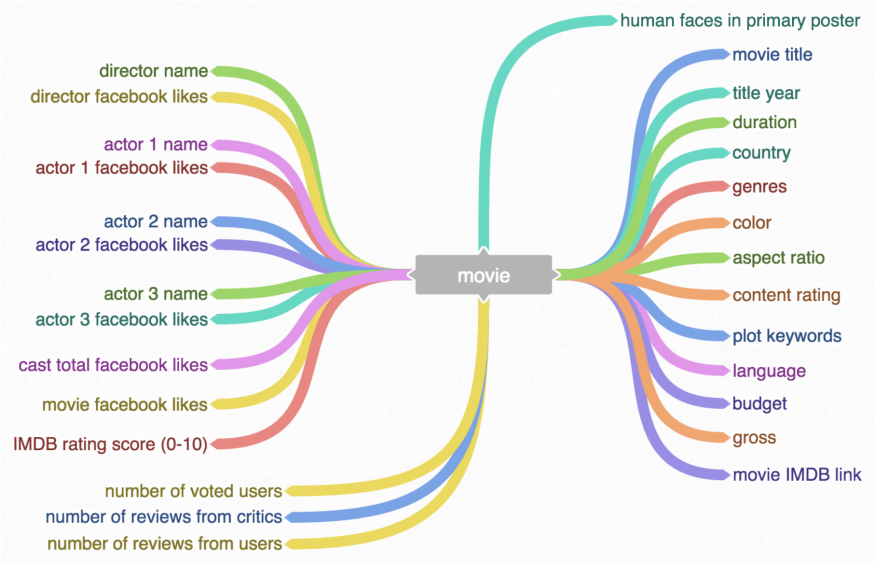

Figure 1: Variables Related to Movie