Predicting Box Office Revenue

IIT2013196,IIT2013128,IIT2013172

5 September 2016

1 Introduction

Film industry is a big business in United States.It is one of the biggest player in the entertainment industry.Predicting the Gross revenue movie would be able to generate could be great help to the producers.Till now there is no computational model that can effectively predict the Gross revenue movie will be able to collect.This depends on a lot of factors like release date,number of available theatre screens,budget of the film, etc.

2 Literature Survey

A Lot of factors that affect the revenue prediction have been studied by different researchers.

Do Critical Reviews Really Matter? As mentioned in [5] even though some reviewers may have different tastes than the people reading their reviews, consumers still read the reviews before going to watch the movie to know whether it is worth their money or not. This tells us that critic ratings do matter in the revenue prediction model.

Forswell in [6] collected public information available on IMDB site for movies after January 1st, 1990 and only included those having total box office revenue greater than 100,000 dollars totaling 2500 movies.He then used linear regression model using features first week end revenue,budget and number of available theatre screens.It did not give very good result and when he tried splitting the dataset, number of training samples decreased and performance was reduced.

Robert in [8] divided the set of features in simple , complex and sentiment where simple is numeric only, complex is numeric and text based and sentiment includes all.He thought that this problem can be modeled both as regression and classification. He used logistic regression for classification by making classes by dividing the range of min and max value of gross revenue into buckets.After doing more analysis he found that text and sentiment features were insufficient to predict the revenue.

Some researchers like in [3] tried to predict the revenue before release or after 1st week of revenue. They analysed different features like production budget of the film,revenue of the 1st weekend , sequel movie or not,star power,MPAA(Motion

Picture Association of America film rating system) rating etc. He analysed the change in gross revenue with respect to rating of the movie.He tried to give weights to these categorical variables and used linear regression to do the prediction.

Vitelli in [10] tried to create a set of features and did extract values from graphical properties of the actor-actor, actor-movie, and movie-movie relationships.

Some tried to predict the revenue before the release and some after 1st week of release.Budget of the film seems to different effects on different genre movies. Mostly big budget films which are Action,Animation, Adventure tend to attract a large amount of audience whereas on the other hand movies which even after spending a lot of money like which are biography,drama,etc tend to earn less.This fact has been supported in [4].

Genre and revenue analysis was done. In Anast [1] whose tried to showed some negative relationship between the action genre and revenue. Prag and Casavant [2] showed a negative relationship for drama and revenue.

Dursun Delen in [7] used Neural Network with features like MPAA rating,genre,star value,sequel,special value but for prediction of pre-release revenue. It was used because neural networks can handle a mix of continuous and discrete values pretty well.

Till now all the researchers were considering all the different factors that affect the movie success.But Thorsten in [9] did a study on inter dependence between these features whether success or effect of one feature can affect the effect of other on revenue.For example,does advertising influence the box office revenue directly by creating a media presence, or indirectly through impacting consumers quality perceptions of the movie?

Overall all the researchers tried to estimate the Gross Box office revenue movie will be able to collect.Some considered it to be regression problem and few as classification problem.None of them tried to analyse it genre specific. As if saying that all these features like MPAA rating , production budget of the film, marketing budget of the film,critic rating in the first week are not all important for all genre. Biography or Drama movie may be good spending a lot of money but would not be able to generate a lot of revenue.Likewise, only action and animation big Budget films attract a lot of audience not true for other. Some superhero movies like Super Man and others saying Big Action Movies people don't care about critic ratings they just see it. We should not consider Critic Rating in estimating Action movie revenue. And , budget of the film should not a parameter in Drama or Biography films. Talking about MPAA ratings, movies that are PG-13 , R and action,adventure,horror affect the revenue. They should be considered as a parameter. Whereas movies made for children like G rated , they when released in US mostly collect almost equal amount of revenue.

3 Table

Name	Information
Algorithm and tools	Regression, Classification, Neural Network.
Set of features	movie rating, critic views, theatre screens,budget,release date.
DataSet	Rotten Tomato, Wikipedia, IMDB.
Genre	Sci-Fi, history, thriller,horror, comedy, cartoon, action, documentary

4 Our Approach

1. We want to estimate Domestic gross revenue of movie after 1 week of release (Post Release).
2. We will collect public information available about movies from websites like Rotten Tomatoes, Wikipedia, IMDB, Box Ofice Mojo, FilmCritic etc. only for English Movies in United States.
3. We will analyse differnt feature relationships on the basis of genre.
4. We will try Regression and classification models on different set of features.
5. We will try different Machine Learning models like Neural Network to enhance accuracy of prediction.

References

[1] *Anast, P. 1967. 'Differential movie appeals as correlates of attendance', Journalism Quarterly.*

[2] *Prag, J.J. Casavant, J. 1994. 'An empirical study of the determinants of revenues and marketing expenditure in the motion picture industry', Journal of Cultural Economics,.*

[3] *Simonoff, J. S. and Sparrow, I. R. Predicting movie grosses: Winners and losers, blockbusters and sleepers. In Chance, 2000.*

[4] Luis Cabral and Gabriel Natividad. *Box-Office Demand: The Importance of Being.* 2013.

[5] Alec Kennedy. *Predicting Success: Do Critical Reviews Really Matter?* 2010.

[6] Anahita Sidhwa Nikhil Apte, Mats Forssell. *Predicting Movie Revenue?* 2011.

[7] Dursun Delen Ramesh Sharda. *Predicting box-office success of motion pictures with neural networks.* 2006.

[8] David Cummings Steven Yoo, Robert Kanter. *Predicting Movie Revenue from IMDb Data*. 2011.

[9] Gianfranco Walsh Thorsten Hennig-Thurau, Mark B. Houston. *Determinants of Motion Picture Box Office and Profitability: An Interrelationship Approach*. 2008.

[10] Matt Vitelli. *Predicting Box Office Revenue for Movies*. 2012.