

Box Office Prediction

Rahul Paravas 12539

Divyank Aggarwal 13258

Introduction

In this project we aim to predict net box office gross of movies using its pre release data such as genre, actors, director, etc. as well as news analysis. The motivation for using news analysis is that traditional models using only movie data have achieved limited success.

Related works

Early efforts with movie prediction using traditional data have yielded limited success. High success rates have been achieved with news analysis (**Zhang and Skienna, 2009**) as well as with social media such as twitter (**Sitaram and Huberman, 2010**).

Approach

Our project will be divided into two parts – buiding appropriate dataset (traditional data as well as news data) , and regression analysis of the dataset.

Traditional dataset will be sourced from imdb, boxofficeindia.com

News articles will be sourced from popular english news dailies as well as popular internet articles and analysed using Lydia, a high speed text processing system or google news.

The analysis will be done using traditional regression models and k-Nearest neighbours model.

References

- 1) Improving movie gross prediction through news analysis, Wenbin Zhang and Steven Skienna. 2009
- 2) Predicting the future with social media, Sitaram Asur and Bernardo A. Huberman. 2010
- 3) Early prediction of Movie box office success based on Wikipedia activity big data, Márton Mestyan , Taha Yasseri , János Kertész. 2013