VAST 2013 Mini-Challenge 1: Box Office VAST - Team VADER

Yafeng Lu*
Arizona State University

Feng Wang
Arizona State university

Ross Maciejewski, *Member, IEEE*
Arizona State University

ABSTRACT

VAST 2013 Mini-Challenge 1 was a rolling competition in which participants would submit a set of weekly predictions for movie revenue and viewer rating. This was a closed world contest in which contestants were provided with a set of Twitter indices, bitly links, and access to the Internet Movie Database. In order to facilitate these predictions, we have created a web-deployable system that combines predictive techniques (multiple linear regression), data mining (sentiment analysis), and interactive visualizations for predicting the opening weekend gross and viewer rating scores of upcoming movies. Taking use this system with human in the loop, we got 4 times top predictor in 11 entries of submission.

Keywords: Multivarite regression, Box office prediction, Social media data, Twitter, Visual analytics

1 INTRODUCTION

Movie revenue prediction has drawn the most attention in both movie industry and academic field. Movies meta-data, social media data, text extraction and analysis, and google search volume have been explored in many prediction methods. In this work, we explore augmenting traditional box office modeling techniques through the use of visual analytics.

2 MULTIPLE LINEAR REGRESSION MODELING

Regression analysis is one of the most widely used methods for pattern detection, and a substantial body of literature exists on developing multiple linear regression models [2] for movie revenue prediction (e.g., [4, 5]). Traditional variables used in these box office prediction models include known variables (e.g., the MPAA rating, the number of screens) and derived measures (e.g., popularity of the movie stars, popular sentiment regarding to the movie).

Based on our initial literature search, we chose to utilize multiple linear regression for an initial prediction range for the opening weekend box office revenue. Due to the closed world nature of the contest, traditional variables used by other researchers were not always available (for example, theater count is not provided for every movie in IMDB). As such, we explored a variety of different variables available in the contest, see Table 1.

After initial model fitting using R [3], we found the best model to predict the opening weekend gross from our variables was:

$$OW = \beta_0 + \beta_1 \text{TBD} + \beta_2 \text{Budget}.$$

Parameters are fit using movie data released beginning in January of 2013. Our first submitted prediction was for the May 17th weekend and used data from 39 movies for training. Each week model parameters were updated with the newly collected data, and our weekly models reported an *adjusted* $-R^2 \approx 0.60$ with $p < .05$. Our final parameters were $\beta_0 \approx 4.9 \times 10^3$, $\beta_1 \approx 4462$, and $\beta_2 \approx .23$.

---

*email: lyafeng,fwang49,rmacieje@asu.edu

Variable	Description
OW	3-day Opening Weekend Gross
Budget	Approximate movie budget from IMDB
TBD	The average daily number of tweets over the 2 weeks prior to release
TSS	Tweet Sentiment Score - A summation of each individual word's sentiment polarity as calculated via SentiWordNet [1]
MSS	Movie Sentiment Score - A derivation of the overall sentiment of a movie
MSP	Movie Star Power - A summation of the twitter followers of the three highest billed movie stars (as listed by IMDB)

3 TWEET SENTIMENT MINING

While multiple regression was found to be a reasonable starting point for predicting the box office gross, predicting the review score required text analysis. We wanted to approximate tweet sentiment as it relates to upcoming movies. First, each tweet is processed according to the SentiWordNet dictionary [1], where each word in the tweet is assigned a score from -1 to 1, where -1 is the highest negative sentiment score and 1 is the highest positive sentiment score. Next, each tweet is assigned a sentiment score by summing the sentiment score of all words in the tweet and normalizing the range from $-.5$ to .5 (TSS in Table 1). Finally, the movie sentiment score (MSS in Table 1) is calculated as

$$MSS = \frac{Positive\ Score}{Positive\ Score + Negative\ Score}$$

where *Positive Score* is the sum of all tweets for a given movie with a TSS greater than zero and *Negative Score* is the absolute value of the sum of all tweets for a given movie with a TSS less than zero.

4 VISUALIZATION TOOLS FOR FILTERING AND ANALYSIS

Part of the challenge in utilizing twitter and bitly data for prediction is the noise inherent in the data. In order to deal with this challenge, we have developed a variety of visualizations that allow for interactive analysis and filtering for noise reduction.

4.1 Tweet Bubble Chart

For sentiment analysis, some tweets follow a pattern of "I want to see this so bad." In this example, 'bad' will be classified as a negative sentiment word with the rest being neutral; however, this statement is actually quite positive about the movie. Thus, it is clear that a completely automatic analysis of tweet sentiment could result in large errors. To compensate for this, we have developed an interactive bubble chart, Figure 1, which allows users to modify the sentiment of any given tweet.

In Figure 1, the user has moused over a tweet and sees that the user is wondering if the movie will be good. From there, the user can interactively set this sentiment to neutral, which will update the database and all associated models using that variable. As part of the bubble chart widget, users can filter by keyword, thus in the example of "I want to see this so bad," if there is a common keyword that is causing sentiment misclassification, the user can quickly adjust the overall sentiment score. Tweets can also be filtered based on their overall sentiment score using the legend.
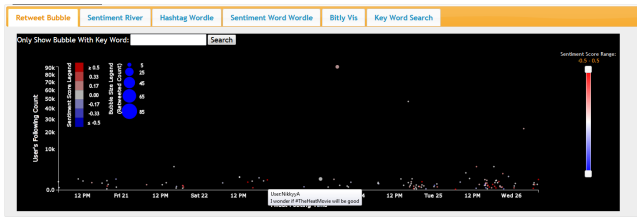
Figure 1: Our sentiment tweet bubble chart for The Heat. Each circle is a tweet. Red represents positive sentiment, blue negative. The size of the bubble represents the number of times a tweet has been retweeted, the x-axis is time, and the y-axis is the number of followers that the user who submitted the tweet has.

Along with sentiment noise, there is also noise in the underlying data collection itself. A specific example of this would be the movie, "The Heat". This movie shares its name with the popular professional basketball team, The Miami Heat. To further compound issues, this movie was released near the same time that the Miami Heat won the NBA championship, resulting in a large number of tweets referencing "The Heat" being unrelated to the movie. The bubble chart tool can also be used to mark tweets for removal from the dataset by right clicking on a tweet, thus updating the TBD variables from Table 1 which in turn updates the regression model. A list view of all modified tweets is also available to allow collaborative analysis in which multiple analysts can see modifications and adjust items based on their own expertise.

4.2   Similarity Visualization

The most used feature of our system was interactive similarity matching. The goal was to enable users to see which past movies are similar to the current film under prediction, then we can see if our model typically underestimates, overestimates or is relatively accurate. This allows us to refine our final prediction value for both the box office gross and the review score.

We have nine similarity criteria: TBD, tweet trend, tweet sentiment, MSS, MPAA rating, genre, MSP, sentiment wordle and predicted gross. The distance between two movies, when using any similarity criteria, is calculated using a Euclidean distance metric. In terms of categorical similarity (such as MPAA), we show the most recently released movies that have the same MPAA rating. In all similarity matches, we show the top five most similar movies (Figure 2 is cropped to show the top-most similar movie).

5   VADER PREDICTION RESULTS

From May 17th through July 26th, we predicted the box office gross and review score for 23 movies (only 21 are shown in the chart as two movies (The Bling Ring and The To Do List) were limited release movies). Our typical process began with utilizing the bubble chart to adjust the data for noise and sentiment values as a means of data cleaning for the regression model. Next, the bitly widget was used to extract review scores from web articles. The average review score was analyzed and a mental model for the review score prediction was developed. Subsequently, the similarity widget was used to compare previous opening weekend predictions and review scores of similar movies. Finally, the opening weekend gross predictions were then summed and the total was compared to the total weekend prediction for final adjustments.

Average errors for all predictions are given in Table 2. Note that we provide the average error for using our regression model with no modification in Table 2 and we see that it is nearly double that of our final prediction informed through visual analytics. After all, the model only explains 60% of the variance of the data. Overall, our prediction error (.281) was only a slightly higher margin of
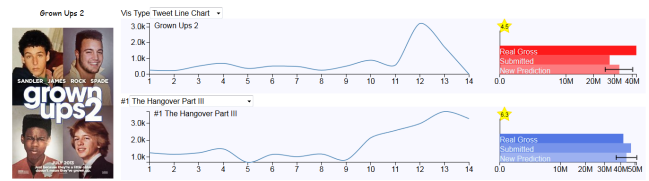

Figure 2: A user defined similarity view cropped to show the top-most similar movie. In the center is the tweets by day view, on the right is a graph of the opening weekend gross. There are bars for the actual gross, our final prediction, and the prediction range. The star in the upper left corner of the graph shows the review score.

Prediction Source	Average MRAE
Our Submitted OW Prediction	0.281
Regression Model OW prediction	0.447
boxoffice OW Prediction	0.183
filmgo OW Prediction	0.265
hsx OW Prediction	0.205
boxofficemojo OW Prediction	0.230
Our Submitted Viewer Rating prediction	0.078

error than filmgo (.265), but approximately 50% worse than boxoffice.com (.183). However, if we remove the After Earth and Now You See Me weekend which with prediction uses only the model, our MRAE drops to .239 which puts us solidly in the prediction range of several experts (HSX and boxofficemojo). Other sources of error can be accounted for in disrupted twitter and bitly data feeds. These interruptions were pronounced for The Heat, White House Down, Monsters University and World War Z. However, even with those interruptions, our predictive analysis process was still quite robust.

6   SUMMARY

It is clear that visual analytics can be used to improve predictive models by bringing domain knowledge into play. While our systems visuals are relatively traditional, the combination of these with analytical methods has proven very effective. Of the 11 weeks in which contest entries were submitted, our team was the top predictor four of the weeks, our predictions had a lower MRAE than the pros in week 2 (as well as for multiple movies over the course of the contest, and we were consistently one of the top 3 teams for overall prediction. Currently no other team has been the top predictor more than twice. For a demonstration of the full system, please see the accompanying video at `http://www.youtube.com/watch?v=dQuti7aHvIw&fmt=22`.

REFERENCES

[1] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the International Conference on Language Resources and Evaluation*, 2010.

[2] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to Linear Regression Analysis*. Wiley, 2012.

[3] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, 2008.

[4] A. C. Reggie Panaligan. Quantifying movie magic with google search. *Google Whitepaper — Industry Perspectives + User Insights*, 2013.

[5] J. S. Simonoff and I. R. Sparrow. Predicting movie grosses: Winners and losers, blockbusters and sleepers. *Chance*, 13(3):15–24, 2000.