OPENING BIG IN BOX OFFICE? TRAILER CONTENT CAN HELP

Adarsh Tadimari [1], *Naveen Kumar* [2], *Tanaya Guha* [3], *Shrikanth S. Narayanan* [2]

[1]Electrical Engineering, Indian Institute of Technology Madras, Chennai, India.
[2] Signal Analysis and Interpretation Lab, University of Southern California, Los Angeles.
[3]Electrical Engineering, Indian Institute of Technology Kanpur, India.

ABSTRACT

Computational prediction of a movie's financial success usually relies *only* on metadata such as - genre, budget, actors, Motion Picture Association of America (MPAA) rating and critics' reviews. We argue that movie trailers, created to invoke viewers' interest and curiosity about a movie, carry complementary information for predicting a movie's financial future. We created a database consisting of 474 American movie trailers along with various metadata and information about movie's financial success in the opening weekend. A number of features that capture the emotional information contained in the audiovisual stream of the trailers are designed and extracted. We observe that the content-based features have as much predictive information as the meta features. Through regression analysis on our database, we show that signal information from trailer content improves the prediction performance.

Index Terms— Content analysis, movies trailers, multimedia, success prediction.

1. INTRODUCTION

The first weekend after a movie's release is crucial for economic success. A movie can collect as much as a quarter of their entire box office gross in the opening weekend [1]. Strong opening weekend sales also sets about international ticket sales, home entertainment, DVD sales, TV and the cable markets. Films that falter during the opening weekend often fail to attract people thereafter, and are thus less successful.

To maximize revenue in early stages, filmmakers invest heavily in promoting their movies through print and web media, social networks and TV before release [2]. The primary promotion content for a movie is its trailer. According to the Motion Picture Association of America (MPAA), filmmakers allocate at least 4% of their marketing budget on theatrical trailers [3]. A movie trailer is created by combining exciting and noteworthy parts of the movie to invoke viewer interest and curiosity. In this work, we investigate the relationship between movie trailers and their opening weekend box office collection.

The idea of developing computational models to predict the financial success of a movie is not completely new. The majority of existing literature however, relies on metadata (genre, budget, actors, etc.) to build the prediction system [4, 5]. While one study chooses to use critic's reviews [6] to predict a movie's success, others using metadata descriptors, such as, the release period of the movie (e.g. summer, christmas), MPAA ratings, actor, budget, sequel, to predict the opening weekend gross along with the total gross of the movie and its length of run showed that the budget, genre and MPAA rating are significant in predicting a movie's success [4]. In addition, they showed that it is easier to estimate the initial performance of the movie than the total performance since the latter depends on various additional factors such as competitive force. There has been work analyzing the impact of star power on the success of the movie showing that stars could affect the odds of a movie's success [7, 8]. Twitter chatter and Google search trends before and after the release of a movie have also been used to predict box office success [2]. Positive references to a movie in weblogs has been shown to correlate with the movie's financial success [9]. Activity levels of viewers and editors of movie entries in Wikipedia has been used to predict the popularity of the movie [10]. Quantitative news data has been shown to be helpful in predicting movie grosses [11].

This work aims at investigating the impact of the *content* of a movie's trailer on its initial financial success. This differs from the most of the existing work that uses only metadata for prediction. Our goal is to predict the opening weekend's collection of a movie using audiovisual features extracted from the trailer content, along with metadata. Our contributions are two-fold: (i) Creation of a database containing trailers of 474 American movies along with various metadata (budget, genre, etc) associated with the movies. The information about the opening weekend's gross is collected from IMDb (see section 2 for details) (ii) Designing a set of low-level audiovisual features that can be extracted from the trailer content. Through various experiments, these features are shown to improve meta-data-only prediction systems.

2. OUR DATABASE

To study the impact of trailer content on a movie's initial success, we create a database consisting of movie trailers, and various metadata associated with them. The success metric that we choose is a movie's box office collection in the first weekend. Details about this newly created database are as follows.

2.1. Movie trailers

We collect trailers[1] for 474 American movies, released during the years of 2010 to 2014. For each movie, we use only a single trailer, even though multiple trailers may be available. Our database contains trailers of 111 movies from 2010, 104 from 2011, 102 from 2012, 80 from 2013 and 77 from 2014. These are the movies for which the estimated budget and the opening weekend sales information is available on the Internet Movie Database (IMDb).

2.2. Metadata

We obtain the following metadata for the 474 movies from IMDb. This set of metadata was collected, following a recent work [4],

[1]our database of movie trailers along with the metadata is available at *https://github.com/tadarsh/movie-trailers-dataset*

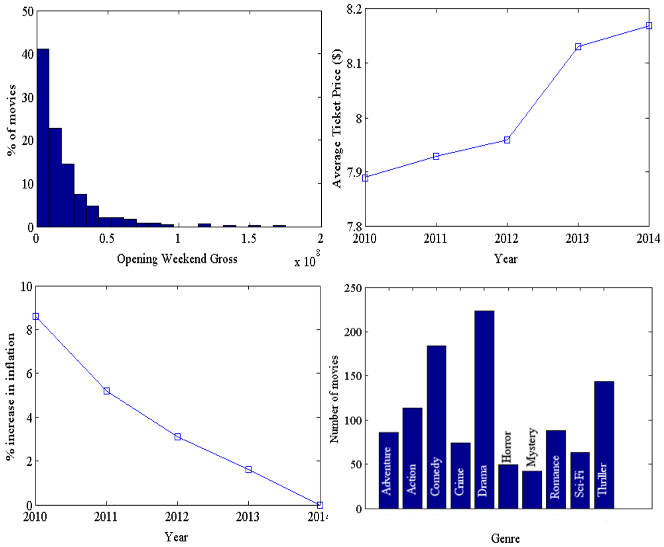

Fig. 1. (Top Left) Distribution of the adjusted opening weekend gross in the obtained data. (Top Right) Average ticket price in US. The data was obtained from NATO (National Association of Theater Owners). (Bottom Left) The percentage inflation indicates the percentage increase in price from a particular year to 2014. (Bottom Right) Distribution of movies across different genre in our sample

where they have been shown to be useful in predicting a movie's success.

- **Production budget:** A movie's budget is usually confidential, hence its estimated budget information is collected and used in our work. We also adjust for the inflation in costs over the past 5 years. Fig. 1 shows the percentage increase in costs from a given year to 2014. This information is calculated based on the consumer price index published by Bureau of Labor Statistics (BLS). The budget information of all movies is adjusted to their inflated cost in 2014.
- **Genre:** The following genre labels are considered in our database: adventure, action, comedy, crime, drama, horror, mystery, romance, sci-fi and thriller. These genre labels are the most common labels in the entire IMDb database. A movie may belong to one or more genres. Fig. 1 (bottom right) shows the distribution of movies across different genres.
- **MPAA ratings** determine a film's suitability for certain section of audience based on its content. Our set of movies has the MPAA ratings of G, PG, PG-13 and R.
- **Release period:** We use four labels: Summer (May-August), Christmas (November and December), Easter (March and April) and Other.
- **First-week screens:** This indicates the number of screens that showed the movie in the opening week.
- **Sequel:** This metadata feature indicates whether a movie is a sequel or not. This is manually annotated.
- **Actor's experience:** We code the number of movies the main actor has appeared prior to the release of a particular movie.

Fig. 2. Trailer of *The Beaver* (left) has the highest amount of blue content (hue of 180-210 on a scale of 360), trailer of *Moonrise Kingdom* (right) has the highest amount of orange content (hue of 30-60 on a scale of 360)

2.3. Opening Weekend Gross

We predict a movie's economic success in terms of the opening weekends ticket sales. We obtain the information about opening weekend gross in US for the movies in our database through IMDb. The amount is adjusted to reflect the box office collection a movie would make in 2014 based on the changes in average ticket prices in US (see Fig. 1, top right). Fig. 1(top left) shows the distribution of opening weekend gross in our sample.

3. FEATURE DESIGN

We extract the following visual and audio features from the the trailers in our database with the general objective of capturing emotion-related information in the content.

3.1. Visual Features

Intensity and color: We propose to use intensity to differentiate between the brighter and the darker movies. For a pixel in an RGB image, the intensity I can be computed as $I = (R+G+B)/3$. We find the mean intensity of all pixels in a frame and compute a 5 bin histogram $i = [i_1, i_2, .., i_5]$ over all frames in the movie trailer. We use a small number of bins as we only want to distinguish between bright and dark movies. When the movies are ranked based on the mean intensity over all frames, we find that trailer of *Despicable Me 2* has the highest mean intensity while the trailer of *Annabelle* has the lowest mean intensity. Color also has been shown to influence human emotions [12]. To capture the color distribution in movie trailers, we use hue, which captures the pure color information of a pixel and is independent of its shade. For higher interpretability of the color information, we computed a 12 bin histogram $h = [h_1, h_2, .., h_{12}]$ pixels of all frames in the movie trailer after converting the frames in RGB space to HSV color space. Figure 2 shows trailers which have the most amount of blue and orange content respectively.
Shots: Since all the movie trailers are of similar length, the number of shots s would approximately represent the frequency of cuts in a trailer. A large number of shots in a trailer would indicate that there are frequent cuts which creates a perception of speed to the viewers [13]. We use an open source tool called *ffmpeg* which can identify the shot boundaries in a video. It detects the shot boundaries based on the pixel differences in consecutive frames.
Motion Activity Like cuts, motion activity is also correlated with a sense of pace and is often used in movie content analysis [13]. Motion activity in a scene is understood in terms of the motion of actors, objects or camera in that scene. This is often used to induce a notion of speed to the viewers e.g., in the action scenes of a movie. To compute motion activity, we use an approach similar to [13]. Since the motion activity m_s is computed per shot, the mean, minimum, maximum and the standard deviation of the motion activity in the

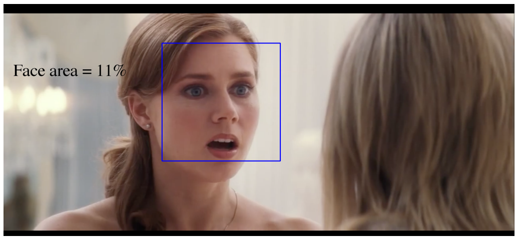

Fig. 3. A sample close-up shot from the trailer of *Leap Year* which has the highest number of close-up shots

LLD (32)	Functionals (12)
ZCR	Mean
RMS - Energy	Standard deviation (std)
F0	Kurtosis, Skewness
HNR	extremes: position, range, value
MFCC 1-12	linear regression: slope, offset, MSE

Table 1. Audio features used in this work: Low level descriptors (LLD) and functionals. ZCR - Zero crossing ratio, RMS - Root mean square, F0- Fundamental frequency, HNR - Harmonic noise ratio, MFCC - Mel frequency cepstral coefficients

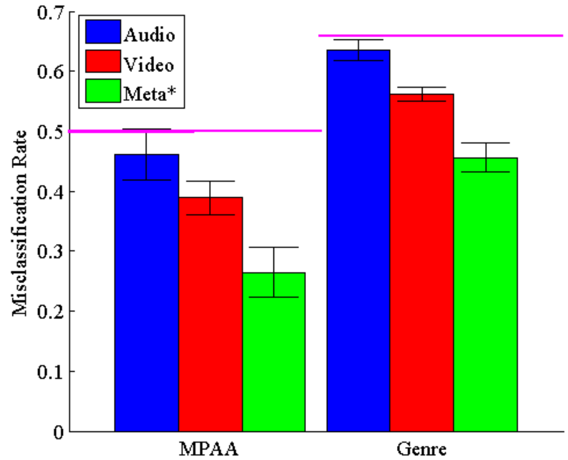

Fig. 4. Results of predicting MPAA rating and Genre using audio, video, and other metadata (Meta*) features. The *pink* baseline represents the misclassification rate of a majority classifier, a classifier which simply predicts the majority class. Meta* indicates metadata features excluding the one being predicted

shots of the movie trailer are used as features. It is represented as $m = [mean(m_s), min(m_s), max(m_s), std(m_s)]$.

Close up shots: Close-ups shots (as compared to wide shots), in general, are used to show the characters' emotion to the audience [14], and to keep the audience connected to the characters. Figure 3 shows a sample close up shot. We compute the number of shots n_c and the percentage of shots p_c which were close-up shots. Information about close-up shots is represented as $c = [n_c, p_c]$. We label a shot as a close-up shot if it has at least one frame with a human face occupying more than 5% of its area. The visual features are concatenated as $\text{Video} = [i, h, s, c, m]$.

3.2. Audio features

Metadata of a movie captures very little information about the emotion portrayed in the movie, or its trailer. In order to obtain additional information, we use the Interspeech 2009 emotion challenge [15] features extracted using the OpenSmile toolkit [16]. These features are designed to cover the prosodic, spectral and voice quality features, and have been shown to perform well in recognizing emotion in speech [15]. Table 1 shows the included low level descriptors (LLDs). Twelve functionals are computed for the 16 LLDs and their derivatives, leading to $12 \times 16 \times 2 = 384$ features. Principal component analysis (PCA) is used to reduce the dimension of features to 50 ($[a_1, a_2, .., a_{50}]$) preserving 99% of the variance. These features are referred to as the audio features: $\text{Audio} = [a_1, a_2, .., a_{50}]$

4. FEATURE ANALYSIS

Our choice of audiovisual features is such that they are related to certain metadata features. For example, horror and thriller movies are usually darker in visual appearance compared to the movies from other genres; romantic movies may have more close-up shots, while action movies have more motion activity. In addition, horror and thriller movies are more likely to be given an MPAA rating of R. In this section, we investigate the usefulness of our audiovisual features in explaining certain metadata features.

4.1. Genre & MPAA Rating

In order to gauge the predictability of genre from audio and visual features, we consider three categories of genre: (1) *Horror, Mystery, Thriller*, (2) *Drama, Comedy, Romance*, and (3) *Action, Adventure, Sci-Fi*. These categories are found to be the most frequent ones in our database. To ensure equal number of samples in each of the categories, we randomly downsample all classes to be of the minority class's size. We use a linear SVM to predict the category of genre with features from the audiovisual content. All significance tests are performed using a 5×2 cv(cross-validated) paired t-test [17] at 1% level of significance. The error bars in Fig. 4 show the standard deviation of misclassification rate of the 10 different folds in the 5×2 cv test. The *pink* line indicates the misclassification rate of a majority classifier, a classifier which predicts majority class. We find that video and metadata predict genre significantly better than a majority classifier ($p_{\text{meta}} = 8.9 \times 10^{-6}$, $p_{\text{video}} = 1.2 \times 10^{-10}$, $p_{\text{audio}} = 0.124$).

For MPAA ratings, we attempt to classify between two categories, $\{G, PG\}$, $\{R\}$ using a similar approach. We find that using metadata or video modality, our system predicts the ratings significantly better than a majority classifier ($p_{\text{video}} = 3.9 \times 10^{-4}$, $p_{\text{meta}} = 1.5 \times 10^{-5}$, $p_{\text{audio}} = 0.27$).

4.2. Continuous metadata features

To analyze the predictability of continuous metadata features (figure 5), such as, budget, actor experience, and runtime, we binarize the samples using the median as cutoff value. We follow an approach similar to genre and MPAA prediction. All significance tests are performed using a 5×2 cv paired t-test at 1% level of significance. We find that metadata predicts budget, runtime and actor experience significantly better than a majority classifier ($p < 6.6 \times 10^{-7}$ in all the cases). Audio predicts runtime significantly better than a majority

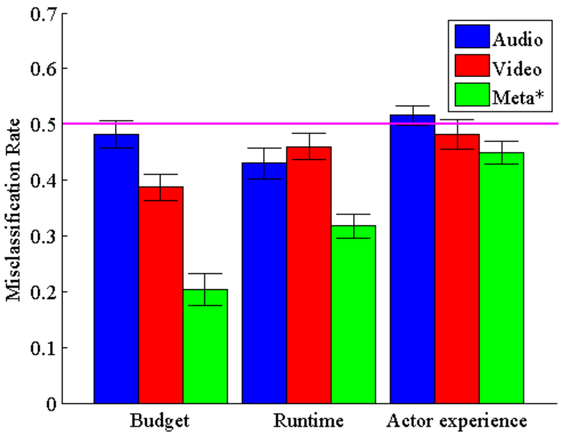

Fig. 5. Results of predicting meta from audio, video, and other metadata (Meta*) features using linear regression analysis. Meta* indicates metadata features excluding the one being predicted

classifier ($p = 1.7 \times 10^{-5}$) while video predicts budget significantly better than a majority classifier ($p = 4.1 \times 10^{-8}$).

We find that the low-level audiovisual features contain information about the genre, MPAA ratings, budget and runtime of the movie but perform poorly in predicting other metadata features (e.g. actor experience).

5. PREDICTING A MOVIE'S INITIAL SUCCESS

A trailer is a primary advertising tool for a movie, and the audiovisual content of the trailer may have a significant impact on the movie's initial success. In this section, we study the correlation between the low level features extracted from the content of the trailer and the opening weekend gross of the movie. We use the adjusted coefficient of determination or adjusted R^2 for evaluating the models.

5.1. Using only metadata

Metadata has been shown to be useful for the task of success prediction [4]. In our experiments, we find that metadata can explain 61% (see Table 2) amount of variance in the opening weekend gross in our database (movies in 2010-2014). This is similar to the result reported in previous work (movies in 2000-2002) [4].

5.2. Using content-based features

In the previous section, we show that some metadata are easily predictable from content-based audiovisual features. This indicates the possibility of using content indirectly to predict the opening weekend gross, i.e. first predicting the metadata from content and then predicting the opening weekend gross. In this section, we investigate the usefulness of the content-based features *directly*. Through regression analysis, we observe that the audio and video features extracted from the content explain 11% of the variance in the opening weekend gross (see Table 2).

5.3. Using metadata and content

On combining the features from metadata and content, we note that there is an improvement of 6.2% in the variance explained by our

Features	R^2	Adjusted R^2
Metadata	0.631	0.613
Audio + Video (Trailer)	0.247	0.11
Metadata + Trailer	0.722	0.651

Table 2. Results on predicting movies' success

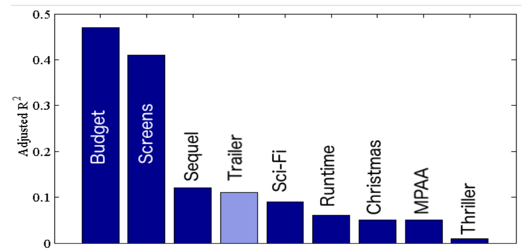

Fig. 6. Comparison of the predictive performance of metadata and the audiovisual features (denoted as Trailer)

model (table 2) when compared to using a metadata-only model. This result suggests that the audiovisual features from the content of the trailer add information complementary to metadata.

Comparison between the model involving metadata and trailer content features, and the one with metadata-only features, shows that the former fits the data significantly better (p-value $= 10^{-3}$) indicating that the content-based features carry additional information. An F-test is used to compare the two models. The F-score is computed such that the number of variables in the two models is accounted for.

$$F = \frac{(SS_1 - SS_2)/(df_1 - df_2)}{SS_2/df_2}$$

where SS_1, SS_2 are the residuals of the two models and df_1, df_2 represent the degrees of freedom (number of variables - 1) of the two models. Figure 6 compares the predictive power of metadata and audiovisual features. In addition to this, we also look at the movies which deviated the most from the model. The top five outliers are *Iron Man 3, The Hunger Games, Alice in Wonderland, Fury, John Carter*. Three of these 5 movies are among the highest grossing movies of all time. This might be explained by the hype created about the movie (an important factor in making a successful opening for a movie) before its release which is not captured by any of our features, the content or metadata. Adding features which account for such factors like hype may help us improve the predictive power further [18].

6. CONCLUSION

In this work, we have shown that taking into account the content information (largely ignored by past studies) present in movie trailers, can improve the current state-of-the-art in success prediction of movies. We have designed a number of audiovisual features, and have shown that these features (especially the visual features) are as powerful as some of the metadata, such as genre and MPAA rating. For the purpose of conducting this study, we have created a database of more than 400 movie trailers and associated metadata, which we have made available for public. As shown in this paper, the low-level audiovisual features are predictive of a movie's genre. Thus our proposed model may be applicable to genre identification. Mining interesting features from social media that capture the hype or awareness of general population about a movie may also be a promising direction of future work.

7. REFERENCES

[1] Jeffrey S Simonoff and Ilana R Sparrow, "Predicting movie grosses: Winners and losers, blockbusters and sleepers," *Chance*, vol. 13, no. 3, pp. 15–24, 2000.

[2] S. Asur and B.A. Huberman, "Predicting the future with social media," in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, Aug 2010, vol. 1, pp. 492–499.

[3] Dave Roos, "How movie marketing works," Sept. 2008, Available at `http://entertainment.howstuffworks.com/movie-marketing.htm`.

[4] Byeng-Hee Chang and Eyun-Jung Ki, "Devising a practical model for predicting theatrical movie success: Focusing on the experience good property," *Journal of Media Economics*, vol. 18, no. 4, pp. 247–269, 2005.

[5] Ramesh Sharda and Dursun Delen, "Predicting box-office success of motion pictures with neural networks," *Expert Systems with Applications*, vol. 30, no. 2, pp. 243–254, 2006.

[6] Mahesh Joshi, Dipanjan Das, Kevin Gimpel, and Noah A. Smith, "Movie reviews and revenues: An experiment in text regression," in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Stroudsburg, PA, USA, 2010, HLT '10, pp. 293–296, Association for Computational Linguistics.

[7] W Timothy Wallace, Alan Seigerman, and Morris B Holbrook, "The role of actors and actresses in the success of films: How much is a movie star worth?," *Journal of cultural economics*, vol. 17, no. 1, pp. 1–27, 1993.

[8] Anita Elberse, "The power of stars: Do star actors drive the success of movies?," *Journal of Marketing*, vol. 71, no. 4, pp. 102–120, 2007.

[9] Gilad Mishne and Natalie S Glance, "Predicting movie sales from blogger sentiment.," in *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, 2006, pp. 155–158.

[10] Márton Mestyán, Taha Yasseri, and János Kertész, "Early prediction of movie box office success based on wikipedia activity big data," *PloS one*, vol. 8, no. 8, pp. e71226, 2013.

[11] Wenbin Zhang and Steven Skiena, "Improving movie gross prediction through news analysis," in *Proc. Int. Joint Conf. Web Intelligence and Intelligent Agent Technology*. IEEE Computer Society, 2009, pp. 301–304.

[12] KAYA NAz and Helena Epps, "Relationship between color and emotion: A study of college students," *College Student J*, vol. 38, no. 3, pp. 396, 2004.

[13] Tanaya Guha, Naveen Kumar, Shrikanth S Narayanan, and Stacy L Smith, "Computationally deconstructing movie narratives: An informatics approach," in *Proc. of the ICASSP 2015*, 2015, pp. 2264 – 2268.

[14] Mary Ann Doane, "The close-up: scale and detail in the cinema," *Differences: A Journal of Feminist Cultural Studies*, vol. 14, no. 3, pp. 89–111, 2003.

[15] Björn Schuller, Stefan Steidl, and Anton Batliner, "The interspeech 2009 emotion challenge.," 2009.

[16] Florian Eyben, Martin Wöllmer, and Björn Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the international conference on Multimedia*. ACM, 2010, pp. 1459–1462.

[17] Thomas G Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural computation*, vol. 10, no. 7, pp. 1895–1923, 1998.

[18] Reggie Panaligan and Andrea Chen, "Quantifying movie magic with google search," *Google Whitepaper Industry Perspectives+ User Insights*, 2013.