# MOVIE TRAILERS & THEIR IMPACT ON SOCIETY

*A Project Report*

*submitted by*

**ADARSH A TADIMARI**

*in partial fulfilment of the requirements*

*for the award of the degree of*

**BACHELOR OF TECHNOLOGY**

*under the guidance of*

**Prof. S Umesh**

**DEPARTMENT OF ELECTRICAL ENGINEERING**

**INDIAN INSTITUTE OF TECHNOLOGY, MADRAS.**

**MAY 2016**

# THESIS CERTIFICATE

This is to certify that the thesis titled **MOVIE TRAILERS & THEIR IMPACT ON SOCIETY**, submitted by **ADARSH A TADIMARI (EE12B003)**, to the Indian Institute of Technology, Madras, for the award of the degree of **BACHELOR OF TECHNOLOGY**, is a bona fide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

**Prof. S. Umesh**
Professor
Dept. of Electrical Engineering
IIT-Madras, 600 036

Place: Chennai

Date:

# ACKNOWLEDGEMENTS

I would like to thank Prof. S Umesh for the guidance, support and encouragement he has provided over the past three years. I am deeply grateful to Naveen Kumar, Prof. Tanaya Guha and Prof. Shrikanth S Narayanan who have helped me and supported me through all the projects which we have collaborated on.

Special thanks to my seniors Hariprasad PS, Nikhil Naphade, Abhinav Garlapati, Kishore Natarajan, Prasanna S and Vignesh Krishnakumar. I have greatly benefited from my discussions with them. I thank HyperVerge, a startup which I have been a part of, for permission to use GPU computing resources from time to time.

# ABSTRACT

KEYWORDS: Movie trailers ; Movie success; Gender bias; Media analysis; Gender recognition

Movie trailers are the primary promotional content for movies and are designed to invoke viewers' interest and curiosity about a movie and use several tactics to persuade the audience. Filmmakers invest heavily in design and distribution of movie trailers. In our work, we analyze the impact of content of movie's trailer on its initial success. We design audiovisual features with the goal of capturing emotional attributes in a movie's trailer and find that these features contain significant information in prediction of a movie's success. This is different from most of the previous work in success prediction which has relied on metadata (actor, director, etc) and the social media.

Gender representation in movie trailers and its impact on a movie's success is also studied. There is ample support for the claim that there exists gender bias in Hollywood movies. We investigate the claim in the case of movie trailers and find that there exists no apparent gender bias. We also propose a clustering based gender classification system which provides a significant improvement in gender classification accuracy.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

Movie trailers are the primary means of advertisement for a movie [17, 31]. They are short video clips (restricted to two and a half minutes as prescribed by the National Association of Theatre Owners) and are created by combining the humorous, thrilling and noteworthy scenes in a movie [17]. They are designed to invoke curiosity and interest in the viewers and use several tactics to persuade the audience. Trailers capture the essence of a movie without revealing everything about it. Most people make a decision of watching a movie by looking at its trailer [13]. According to the Motion Pictures Association of America (MPAA), more than 4% of the marketing budget is allocated for theatrical trailers [27]. In certain cases, the task of designing a movie trailer is delegated to companies which specialise in it. The trailers are shown to viewers in the form of television advertisements, previews in theatres and online video content platforms like YouTube, Vine, etc.

In our work, we analyze the impact of movie trailers on the success of a movie. More specifically, we look at the opening weekend gross of a movie which reflects the initial success of a movie. The weekend after the movie's release is critical for economic success of a movie [36]. Not only is it important for drawing attention to the film, as much as one fourth of the box office collection is made during the first weekend [33]. In addition, it is noticed that films that falter during the opening weekend fail to attract audience and are likely to be less successful.

Previous work in movie success prediction has relied on metadata (genre, budget, actors, etc.), trends in social media and critics reviews. Metadata such as MPAA rating, actor, budget, sequel, release period (summer, christmas, etc), genre have been shown to be significant attributes in prediction of movie's success [5, 32]. Trends in social media have also been explored in the prediction of a movie's success. Twitter chatter has been shown to be in useful in movie success prediction [2]. Google search trends before a movie's release are shown to correlate with movie's success [23]. Activity of editors on a corresponding entry in Wikipedia prior to the release of a movie is also indicative of movie's success [20]. Sentiment in blog posts about a movie have also shown to be indicative of a successful movie [21, 28]. The relation between movie reviews and a movie's success has also been studied in detail [3, 15, 26].

In contrast to previous work, we look into the audiovisual content of movie trailers for prediction. We design content based audiovisual features which could capture information about the emotional characteristics of a trailer. We find that these features have as much information as some of the metadata features. Through regression analysis, we show that these features can improve the metadata-only prediction systems significantly.

We also study the gender representation in movie trailers and its impact on a movie's success. Hollywood is believed to be biased towards men. A recent study conducted by the Geena Davis Institute showed that in the top 500 films between 2007 and 2012 only 30% of the speaking characters were female [34]. They also reported that for every female character, there exist 2.25 male characters. In our work, we analyze the onscreen time of male and female characters in movie trailers. Similar study was done in the case of full length Hollywood movies. On analyzing a small set of 17 full length hollywood movie, they found that 36.2% of the screen time was occupied by female characters [11]. We investigate into the apparent gender bias in movie trailers and check if it exists.

We use face detection and gender classification to find the onscreen time of male and female characters. In addition, we propose a face clustering based approach for gender classification, which improves accuracy significantly. In the case of movie trailers, we find no clear evidence of gender bias. We also analyze the impact of gender bias on the initial success of a movie and find that onscreen gender ratio between male and female characters does not provide significant information in prediction of a movie's success.

# CHAPTER 2

# Datasets

In order to study and investigate the various aspects of a movie trailer, we created a dataset of Hollywood movie trailers and collected metadata (actor, director, etc) corresponding to these movies. We also created a gender dataset for comparing the performance of various gender classification algorithms (described in chapter 4). In this chapter, we describe the creation of the datasets in detail.

## 2.1 Movie trailers dataset

We created a dataset of 474 hollywood movie trailers. These were trailers of movies released between 2010 and 2014. Along with the trailer, we also curated metadata related to the movie. In this section, we describe how the trailers and the associated metadata were collected. The dataset is made available to public [1].

### 2.1.1 Obtaining trailers

We used Wikipedia to obtain a list of all Hollywood Movies between 2010 and 2014. The list was 880 movies long. We used YouTube `www.youtube.com` to download the trailers of these movies by searching for the trailer and downloading the video which YouTube suggests to be most relevant.

### 2.1.2 Metadata

We used IMDb (Internet Movie database) to obtain metadata about the movies. The associated metadata was collected based on previous work [5], we obtained metadata features which were found to be significant in the prediction of movie's success. From IMDb, we could collect metadata features for a subet of 474 movies. The following were the features.

- **Production Budget**. Estimated cost for producing a movie. We adjust for inflation in costs. Figure 2.1 shows the percentage increase in cost from a particular year to 2014. All costs

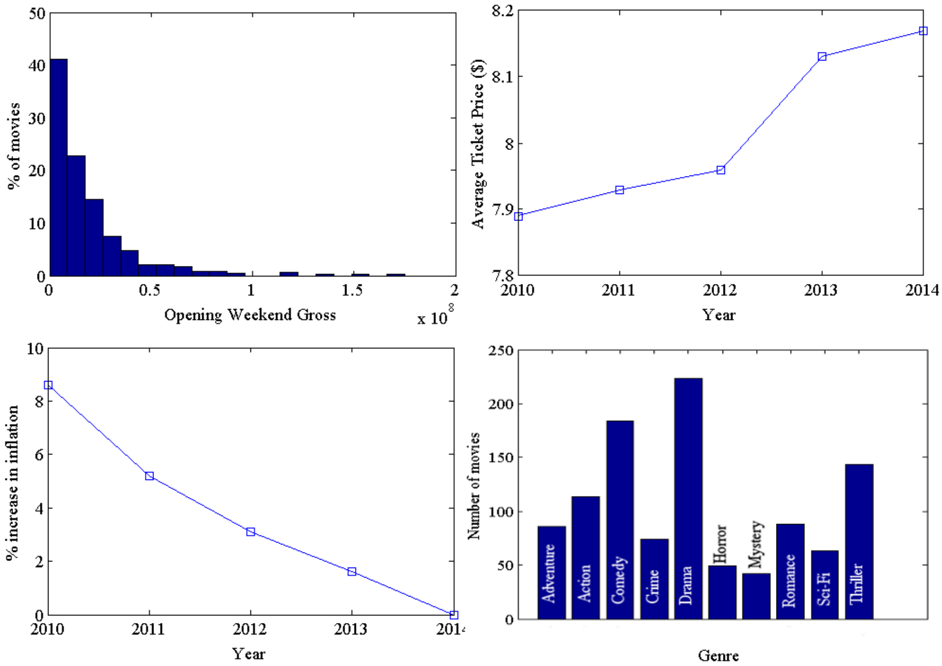


Figure 2.1: (Top Left) Distribution of the adjusted opening weekend gross in the obtained data. (Top Right) Average ticket price in US. The data was obtained from NATO (National Association of Theater Owners). (Bottom Left) The percentage inflation indicates the percentage increase in price from a particular year to 2014. (Bottom Right) Distribution of movies across different genre in our sample

are adjusted to reflect the cost in 2014. For example, if the production budget of a movie in 2010 is $x$ and $y$ is the percentage increase in cost from 2010 to 2014, the adjusted production budget $\hat{x}$ is computed as,

$$\hat{x} = x(1 + \frac{y}{100})$$

- **Genre**. The database has the following genre: action, adventure, comedy, crime, drama, horror, mystery, romance, sci-fi and thriller. These include the most common labels in IMDb dataset. Movie's generally belong to more than one genre. Distribution of genre in the dataset is shown in 2.1.

- **MPAA Ratings**. These ratings determine how suitable a film is for different categories of audience (eg. children). The ratings G (everyone admitted), PG (parental guidance suggested), PG-13 (inappropriate for children under 13) and R (under 17 requires accompanying parent) are seen in our dataset.

- **Release period**. Whether the movie was released in summer (May-August), christmas (November and December), easter (March and April) or other.

- **First week screens**. The number of screens the movie was shown in, in the first week.

- **Sequel**. Whether the movie is a sequel or not. This information was not directly available in IMDb and hence was manually annotated.

| Movie | Male faces | Female faces |
|---|---|---|
| Robocop | 1383 | 1270 |
| Unbroken | 2162 | 884 |
| Big Eyes | 778 | 1504 |
| Dumb and Dumber To | 2291 | 2650 |
| I Frankenstein | 530 | 606 |
| The Imitation Game | 1021 | 850 |
| **Total** | 8165 | 7764 |

Table 2.1: Gender Dataset. The number of male and female faces annotated in each movie in vaidation set.

- **Actor's experience**. The number of movies the main actor has appeared in before the release of a movie.
- **Opening Weekend Gross**. The amount of money the movie made in the first weekend after its release. This is the variable which we are trying to predict. We account for change in average ticket prices to adjust the opening weekend gross. Figure 2.1 shows the change in average ticket prices in the past few years. For example, if the average ticket price of a movie in 2010 is $x$, average ticket price of a movie in 2014 is $y$ and the opening weekend gross in 2010 is $z$. The adjusted gross $\hat{z}$ is calculated as,

$$\hat{z} = z\frac{y}{x}$$

Figure 2.1 also shows how the opening weekend gross is distributed in our dataset.

## 2.2 Gender Dataset

We created a gender dataset which consists of a set of $\tilde{1}6{,}000$ faces labeled as either male or female. Face detection (described in chapter 4) was used to detect all the faces from 6 full length Hollywood movies, Robocop, Unbroken, Big Eyes, Dumb and Dumber To, I Frankenstein, The Imitation Game. The faces were manually annotated as either male or female. The number of male and female faces is balanced in the dataset. Details of the number of faces in each of the movies is shown in table 2.1.

# CHAPTER 3

## Do trailers affect a movie's success?

Movie trailers are the primary promotion content of a movie. In this chapter, we analyze the impact of audiovisual content of a movie trailer on the success of movie. For our analysis, we created a dataset of Hollywood movie trailers. We use linear regression analysis to compare the performance of different sets of features. We also study the effectiveness of various audiovisual features for predicting metadata such as genre and MPAA rating. We find that the audiovisual features contain significant information about a movie's initial success.

### 3.1 Linear Regression Analysis

In order to compare performance among various sets of features, we use linear regression analysis. We fit a linear regression model to the data. We compute a statistic called the coefficient of determination, denoted as $R^2$, which determines the proportion of variance in the dependant variable which is explained by the independent variables. The value of $R^2$ lies between 0 and 1. An $R^2$ of 1 indicates that the regression line fits the data perfectly. The coefficient of determination $R^2$ is computed as follows,

$$R^2 \equiv 1 - \frac{SS_{res}}{SS_{total}}$$

$$SS_{res} = \sum_i (y_i - f_i)^2$$

$$SS_{total} = \sum_i (y_i - \bar{y})^2$$

where $f$ and $y$ represent the predicted and true values respectively. In least squares regression, the value of $R$ weakly increases with the increase in number of independent variables. To account for this inflation the value of $R^2$ is adjusted based on the number of features and samples available for least squares regression. Unlike $R^2$, the value of adjusted $R^2$ denoted as $\hat{R}^2$ improves when new descriptors are added only if they improve $R^2$ more than what is expected by chance.

$$\hat{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

where $n$ and $p$ denote the sample size and the number of independent variables respectively.

## 3.2 Features

We design various audiovisual features in the aim of capturing the emotional attributes of the movie trailer. We describe the video features first and then move on to the audio features.


Figure 3.1: The trailers of the movies *Despicable Me* and *Annabelle* were the brightest and darkest respectively in our dataset whens sorted based on the mean intensity $I$.

### 3.2.1 Video Features

#### Intensity

The intensity of pixel $I$ is defined as, $I = (R+G+B)/3$. We find the mean intensity of all pixels in a frame and then compute a 5 bin histogram $i = [i_1, i_2, i_3, i_4, i_5]$ over all frames in a movie trailer which represents the distribution over number of bright or dark frames in a movie trailer. We use intensity to distinguish between the bright and dark movie trailers. When the movie trailers were sorted based on the mean intensity over all frames, we found that the trailers of the movies *Despicable Me* and *Annabelle* were the brightest and darkest respectively in our dataset (figure 3.1).

#### Shots

Since most of the trailers have similar length, the number of shots would approximately represent the frequency of cuts in the movie trailer [12]. Large number of shots in the trailer would indicate that there are frequent cuts in the trailer. We use an open source tool called ffmpeg which can identify key frames in a video. It detects key frames based on the difference between pixel values in consecutive frames. We use the number of key frames as a measure of number of shots in the trailer.

Figure 3.2: The trailer of Beaver (left) had the highest amount of blue content while the trailer of Moonrise Kingdom (right) had the highest amount of orange content

#### Hue

Color has been shown to influence human emotions [16]. To capture the color distribution in the movie trailers, we use hue, which captures the pure color information of a pixel and is independent of its shade. We computed a 12 bin histogram over pixels of all frames in the movie trailer. Figure 3.2 shows scenes from trailers containing the highest blue and orange content respectively.

#### Motion Activity

To capture the pace of action in the movie trailer, optical flow based motion vector between each pair of frames in a shot were computed based on the standard Lucas-Kanade algorithm. Flow vectors are computed at a large number of key points detected using a corner detection algorithm. Since the motion activity is computed per shot, the mean, minimum, maximum and the standard deviation of the motion activity in the shots of the movie trailer are used as features. We find that the trailer of *Teenage Mutant Ninja Turtles* has the highest mean motion activity.

| 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 3.1: 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

**Close up shots**

Close-ups shots (as compared to wide shots), in general, are used to show the characters' emotion to the audience [8] and keep the audience connected to the movie. In our work, we computed the number of shots and the percentage of shots which were close up shots. We labeled a shot to be a close up shot if it has a frame with a human face occupying more than 5% of its area.

### 3.2.2 Audio Features

We use the Interspeech 2009 Emotion Challenge [30] features extracted using the OpenSmile toolkit [10]. These features were designed to cover the prosodic, spectral and voice quality features. Table 3.1 shows the included LLD's. The 12 functionals were computed for the 16 LLD's and their derivatives, leading to $12 \times 16 \times 2 = 384$ features. Principal component analysis (PCA) was used to reduce the dimension of features to 50 preserving 99% of the variance.

## 3.3 Feature Analysis

In this section, we look at how the hand designed audiovisual features perform in prediction of metadata features. The choice of audiovisual features is such that they are correlated to the metadata. For example, horror movies are darker in intensity and these movies are more likely to be given an MPAA rating of R, action movies have higher motion activity, romantic movies may have more close-up shots.

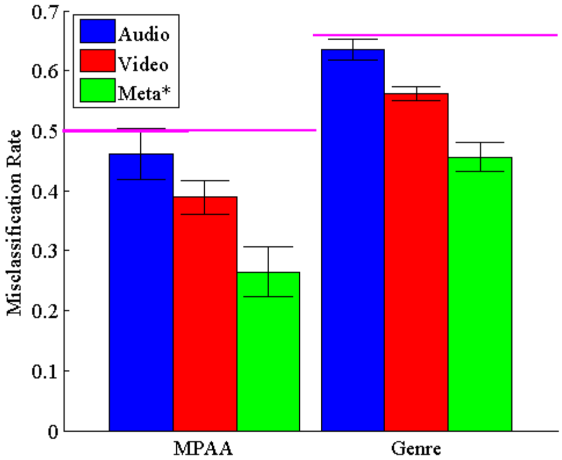


Figure 3.3: Results of predicting MPAA rating and Genre using audio, video, and other metadata (Meta*) features. The *pink* baseline represents the misclassification rate of a majority classifier, a classifier which simply predicts the majority class. Meta* indicates metadata features excluding the one being predicted

### 3.3.1 Genre & MPAA Rating

To compute the prediction performance of genre from audiovisual features, we come up with a 3-class classification problem. We search for frequent itemsets in the dataset and find three supersets of genre: (1) *Horror, Mystery, Thriller*, (2) *Drama, Comedy, Romance*, and (3) *Action, Adventure, Sci-Fi*. We then split the dataset into three categories and also randomly subsample from the class with more number of movies such that all the three categories have an equal number of movies. We end up with 395 movies in total with 95 in each category. We use a linear SVM to predict the category of genre. All significance tests are performed using a $5 \times 2$ cv(cross-validated) paired t-test [7] at 1% level of significance. The error bars in Fig. 3.3 show the standard deviation of misclassification rate of the 10 different folds in the $5 \times 2$ 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 come up with a 2-class classification problem. We consider two supersets, $\{G, PG\}$ , $\{R\}$ and use a similar approach. We find that using metadata or video modality, our system predicts the ratings significantly better than a majority classifier ($p_{\text{video}} = 3.9 \times 10^{-4}$, $p_{\text{meta}} = 1.5 \times 10^{-5}$, $p_{\text{audio}} = 0.27$).

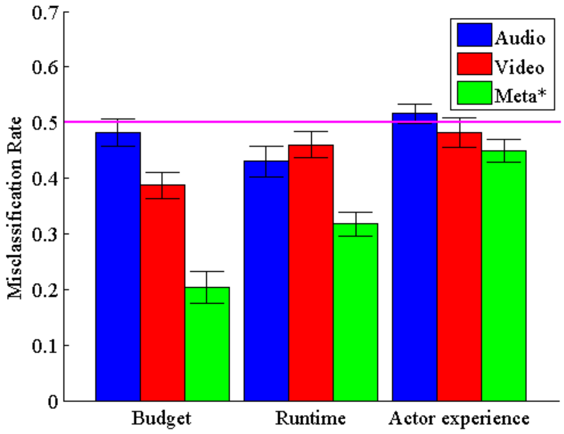


Figure 3.4: Results of predicting meta from audio, video, and other metadata (Meta*) features using linear regression analysis. Meta* indicates metadata features excluding the one being predicted

### 3.3.2 Continuous metadata features

In order to look at how effective the audiovisual features are in prediction of metadata, namely, budget, actor experience and runtime we binarize the values using the mean of each atrribute as the threshold. We follow an approach similar to that used for Genre and MPAA prediction. All significance tests are performed using a $5 \times 2$ 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 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 hand designed audiovisual features contain significant information about the metadata features such as genre, MPAA ratings, budget and runtime of the movie (figure 3.4). Hence, these audiovisual features could be utilized to build systems to predict the metadata features.

## 3.4 Predicting a Movie's initial success

In this section, we study the correlation between the audiovisual features extracted from the movie trailers and the opening weekend gross of the movie. Using linear regression analysis (section 3.1), we show that the audiovisual contain significant information about a movie's success. We

| 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 3.2: Results on predicting movies' success

first compare the results obtained using metadata features with previous work. We then show how addition of content based features can improve the prediction performance.

### 3.4.1 Using only metadata

We try to replicate the performance obtained previously [5]. The previous work was based on movies between 2000-2002 and had reported that the metadata features explain 61.1% of variance in opening weekend gross of a movie. We find that metadata explains a similar fraction of variance in opening weekend gross of a movie (refer to table 3.2).

### 3.4.2 Using content-based features

Predictability of meta data features from audio-visual features indicate that audio-visual features could be useful indirectly for the task of opening weekend gross prediction. For example, we could predict the metadata from the trailer content and then predict the opening weekend gross. In this work, we use the audio-visual features directly. We find that the audio and video features extracted from the content explain 11% of the variance in the opening weekend gross (refer to table 3.2).

### 3.4.3 Using metadata and content

On combining the features from the metadata and content of the trailer, we find that there is a 6.2% improvement in the variance explained by the model (table 3.2) when compared to using a model which has only the metadata features. This suggests that the audiovisual content of the trailer contains complementary information about a movie's success.

### 3.4.4 Complementary information in content features

We compare the model involving metadata-only features and the model involving metadata and trailer content and find that the latter fits data significantly better (p-value $= 10^{-3}$). This indicates that the content-based features carry complementary. An F-test is used to compare the two models.

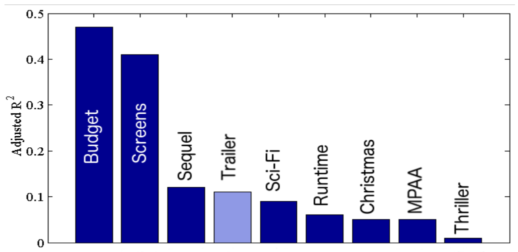


Figure 3.5: Comparison of the predictive performance of metadata and the audiovisual features (denoted as Trailer)

The number of variables in the two models is accounted for when computing the F-score.

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

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 3.5 compares the predictive power of metadata and audiovisual features. It shows that the audiovisual features from movie trailers contain as much information as a few other metadata features (eg. genre, sequel, etc).

## 3.5 Discussion

In this chapter, we looked at the usefulness of audiovisual features from the content of a movie trailer in prediction of a movie's initial success. We find that these features contain significant information about the opening weekend gross of a movie and could be used to improve the prediction performance. We did not study the importance of different audiovisual features. Studying the importance of different features could help in better design of movie trailers to maximize a movie's success in the box office.

The metadata and audiovisual features lack information about how the movie is trending in social media. Looking at the top five outliers to our model in the dataset (*Iron Man 3, The Hunger Games, Alice in Wonderland, Fury, John Carter*), which area also among the highest grossing movies of all time, we find that the hype created about the movie and how the movie is trending is a crucial factor in prediction of a movie's success. Such trends could be explained by other work done in this field such as, looking at the number of tweets before the release of a movie [2] or the number of google searches before a movie's release [23].

## 3.6 Summary

Previous work in the field of movie success prediction has largely ignored the content information from movies. In our work, we look at how the content of a movie's trailer could be used to predict a movie's success and we find that low-level audiovisual features add complementary information (to metadata) in prediction of a movie's success. In addition to predicting success, we find that the audiovisual features designed have significant information about certain metadata features (eg. genre, MPAA) and hence could be used in genre prediction systems.

In the next chapter, we investigate the existence of gender bias in movie trailers. The onscreen time of male and female characters could influence the success of movie. We try to see if such an influence exists.

# CHAPTER 4

# Gender Bias in Movie trailers

Hollywood is claimed to biased towards men. In a study, Gender Bias without Borders [34], conducted by the Geena Davis Institute reported that there in gender inequality in film. They claim that, in hollywood movies, less than 35% of the speaking characters are women. This is alarming considering the fact that half of the movie viewers are female.

In our work, we investigate if there exists any apparent inequality in the onscreen time of male and female characters in movie trailers. There has been similar work done in the case of full length hollywood movies. The female on-screen time averaged over 17 full length hollywood movies was found to be 36.19% [11]. We compute the onscreen female to male ratio over a large set of hollywood movie trailers and also study how it varies with metadata features (genre, MPAA ratings, etc.).

We use a simple appraoch involving face detection and gender classification to compute the onscreen time of characters. This approach has been described in the first section. In addition, we also propose a clustering based voting approach for gender classification. We find that this improves the gender classification accuracy significantly.

## 4.1 Method Overview

In this section, we describe the approach used to compute the onscreen time of male and female characters. This involves two steps. We first detect all the faces in the frames of the movie trailers. We then classify each face as either a male or a female. This gives a count of number of male and female faces in a movie trailer and can be used to get an estimate the total onscreen time of male and female characters. We also propose a clustering based voting method for gender classification.

### 4.1.1 Face Detection

Face detection is a well studied problem in the field of computer vision. Given an arbitrary image, the task is to find the position and size of all faces in the image (figure 4.1). We use the dlib C++

| Algorithm | Gender Accuracy |
|---|---|
| Google SDK | 80.4 |
| CNN* | 71.7 |

Table 4.1: Comparing performance of the Google SDK and CNN based [19] gender classification systems on the benchmark dataset

toolkit (dlib.net) for face detection. We use the default face detection setup which uses Histogram of Oriented gradients [6] as features with Support Vector Machines (SVM) as classifier. We sample 1 frame every 10 frames in a trailer and detect faces in these frames.

Figure 4.1: *Face Detection*. Given an arbitrary image, the task is to find the position and size of all faces in the image.

### 4.1.2 Gender Classification

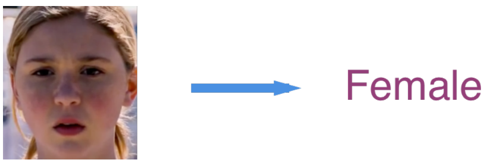


Figure 4.2: *Gender Classification*. Given an image of a face, the task is to classify the face as either a male or a female.

Given the image of a face, the task of a gender classification system is to find the probability of the face being a male or a female face (figure 4.2). Two state-of-the-art gender classification systems were available to us. One of them was a Convolution Neural Network (explained later in the section) model trained for the task of gender classification [19]. The trained model was available online. The other system for gender classification is part of a Google SDK which we were given access to. The specifics of the algorithm are unknown. The performance of the two algorithms on the gender classification dataset described in chapter 2 is shown in table 4.1.

## 4.2 Face Alignment

Faces in movie trailers have different orientations (pose). To extract a feature representation for the face which is invariant to the pose of the face, it is important to perform face alignment. We use Dlib C++ toolkit to detect 5 facial landmarks (figure 4.3). We use OpenCV to perform affine transformation to a fix structure. We fix the size of the facial image to be $144 \times 144$ pixels with a distance of 48 pixels between the eye centers and 48 pixels between midpoint of mouth and midpoint of eyes.

Figure 4.3: *Face Alignment*. We find five facial landmarks and perform an affine transformation to a fixed structure. [37]

## 4.3 Clustering based gender voting

One of the primary contributions of this work is the clustering based voting approach for gender classification. We find that this approach improves performance of existing gender classification algorithms significantly. We first cluster the faces ensuring that each cluster has only faces of one character. We then perform a voting among the cluster to assign a gender for the entire cluster.

In this section, we describe the approach in detail. We first provide the background in Convolutional Neural Networks. We then describe how these networks are used to obtain a representation for faces. We then describe the density based clustering approach used for performing the clustering.

### 4.3.1 Background

The most important step in the process is to be able to cluster faces such that each cluster contains faces of only one person. In this section we describe the basics of Convolutional Neural networks and how these are used to obtain representation for faces.

### Convolutional Neural Networks

In the recent past, Convolutional Neural Networks (CNN) [18] have been used to beat the state of the art in many tasks in the field of image recognition. CNNs are characterized by the convolution operation where the output of a kernel operating on a local window of neurons from one layer is passed on to the next layer. The same kernel is applied throughout the layer, resulting in weights being shared by several connections in the network. The convolution operation can be implemented efficiently on a GPU leading to speedup in training and evaluation. In general, CNN consists of a set of convolution and pooling layers, accompanied in the end by a set of fully connected layers. In the network used in our a work, a max-feature-map layer is used which is introduced in [37].

- **Max-Feature-Map (MFM)** Given a input convolutional layer $C \in R^{h \times w \times 2n}$, MFM computes the following output

$$f_{ij} = max_{1 \leq k \leq n}(C_{ij}^{k}, C_{ij}^{k+n})$$

  where the input convolutional layer has $2n$ channels, and $h, w$ represent the width and height of the convolutional layer input.

### Learning Representation for faces

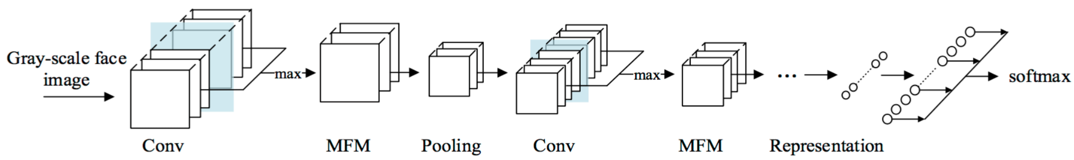


Figure 4.4: Architecture of the CNN model used for extracting representation face feature.

In the past few years, there has been tremendous work in face recognition[22, 24, 29, 35]. Most of the existing state of art algorithms in face recognition use convolutional neural networks to learn robust and discriminative representation. We leverage the existing state of art face recognition models to extract a representation for faces. We use a trained face recognition model [37]. The architecture used in the model is shown in figure 4.4. The representations have been shown to be useful in the task of face verification. This particular model has an accuracy of 97.8% on the verification task. The cosine distance between faces which belong to the same person is lesser than the cosine distance between faces of different people.

**Density Based Clustering**

Figure 4.5: *Face Custering*. Sample clusters obtained on clustering faces using DBSCAN algorithm

Leveraging on the fact that the cosine distance between face representation of faces belonging to the same person is lesser than the cosine distance between faces belonging to different people, our aim is to perform clustering in the space of face representations such that each cluster contains faces of just one person. We use DBSCAN [9], a density based clustering algorithm for the task. We argue that since DBSCAN looks at the distance between two data points to make decisions, it is well suited to the task. The $\epsilon$ and min-pts are tuned on the validation set. The clustering purity was found to be 99.2%. Figure 4.5 shows sample clusters obtained by clustering faces using DBSCAN algorithm.

### 4.3.2 Implementation

We used dlib C++ toolkit to detect the faces in a frame. We used OpenCV [4] for alignment of faces. We used an open-source deep learning toolkit Caffe [14] to extract features for the faces using the

| Algorithm | Accuracy (Individual) | Accuracy (Cluster voting) |
|---|---|---|
| Google SDK | 80.4 | **94.4** |
| CNN* | 71.7 | 82.4 |

Table 4.2: Comparing performance of the Google SDK and CNN based gender classification [19] systems on the benchmark dataset

face recognition model which was made available online [37]. We use Python's Scikit-Learn [25] to perform the density based clustering using the DBSCAN algorithm.

## 4.4 Comparison of various gender classification systems

Table 4.2 shows the performance of various approaches on the validation dataset. We find that the Google SDK along with cluster based voting performs best among all the approaches with an accuracy of 94.4%. We use this system to report results on the dataset of 474 movie trailers.

We find that cluster based voting improves performance of both the systems (Google SDK and CNN) significantly. The appraoch of cluster based voting is universal and is not restricted to gender classification. Prediction of other attributes such as age could benefit from the above setup.

## 4.5 Observations

We use the Google SDK gender classification system along with gender voting to report our results. We compute the onscreen gender ratios in all the available movie trailers. We find that in the dataset of 474 movie trailers, female characters occupy a healthy 56.3% of the screen time. We also look at how the onscreen time ratio varies across various genre and MPAA ratings. Table 4.3 shows how the gender ratio is distributed over genre. We find that movies belonging to the genre of Romance have higher screen time for female characters in their trailers. Surprisingly, horror movies has a high female to male ratio. Trailers of movies belonging to crime and action have a lower female to male onscreen time ratio. Table 4.4 shows the gender ratio is distributed over movies with different MPAA ratings. In this case we find that movies rated PG have a high female to male ratio while the representation of female is lower in the case of movies rate PG-13 and R.

We also study the impact of ratio of screentime of male and female characters in a movie's

| Genre | Female/Male ratio |
|---|---|
| Adventure | 54.5 |
| Action | 41.2 |
| Comedy | 59.5 |
| Crime | **32.9** |
| Drama | 53.7 |
| Horror | **64.7** |
| Mystery | 47.6 |
| Romance | **63.8** |
| Sci-Fi | 48.5 |
| Thriller | 46.5 |

Table 4.3: Variation of ratio of onscreen time of male and female characters across genre.

| MPAA Rating | Female/Male ratio |
|---|---|
| G | 50.7 |
| PG | 61.3 |
| PG-13 | 40.2 |
| R | 38.6 |

Table 4.4: Variation of ratio of onscreen time of male and female characters across movie ratings.

trailer to its opening weekend gross. We find that including the ratio as a feature did not improve the performance significantly.

## 4.6 Conclusion

In this chapter, we looked at the ratio of screentime of female characters to that of male characters in movie trailers of movies released between 2010 and 2014. We find that, on an average, there exists a healthy ratio of female to male screen time in movie trailers. We also note that movies in genre such as romance and horror have a higher female representation in their trailers while movies in genre such as action and crime have a higher male representation in trailers. In our work, we do not account for the roles played by the male or female characters but only account for the total screentime of male and female characters.

On one of main contribution of the work is the cluster based gender voting algorithm. We find that this improves the accuracy of gender classification algorithms significantly. This approach is not restricted to gender classification alone and can be applied to various other attributes which remain constant throughout a video (eg. a person's age in years).

# CHAPTER 5

## Summary

In this work, we studied the influence of trailer on a movie's success. We designed a set of audio-visual features and showed that the content of the movie trailer contains significant complementary information about a movie's success. We also investigated into the screen time occupied by male and female characters in movie trailers and found that, on an average, there exists a healthy ratio in movie trailers in terms of screen time though there exists variation across different types of movie, in terms of genre and the MPAA rating it gets. We also find that the gender ratio in movie trailers does not provide significant information about the opening weekend gross of the movie.

In our work, we have only tried to understand *if* audiovisual features impact a movie's success. Understanding *how* these features influence success would be useful in better design of movie trailers. In the case of gender bias, we have looked at the screentime of male and female characters which depicts only the quantitative aspect of the story. Developing computational models to understand the type of roles played by male and female characters in movies and movie trailer would help uncover gender bias better.

# REFERENCES


[1] **Adarsh Tadimari, N. K. S. N., Tanaya Guha**, Opening big in box office? trailer content can help. *In IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP) 2016*. 2016.

[2] **Asur, S.** 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*, volume 1. IEEE, 2010.

[3] **Basuroy, S.**, **S. Chatterjee**, and **S. A. Ravid** (2003). How critical are critical reviews? the box office effects of film critics, star power, and budgets. *Journal of marketing*, **67**(4), 103–117.

[4] **Bradski, G.** (2000). Opencv : Open source computer vision. *Dr. Dobb's Journal of Software Tools*.

[5] **Chang, B.-H.** and **E.-J. Ki** (2005). Devising a practical model for predicting theatrical movie success: Focusing on the experience good property. *Journal of Media Economics*, **18**(4), 247–269.

[6] **Dalal, N.** and **B. Triggs**, Histograms of oriented gradients for human detection. *In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1. IEEE, 2005.

[7] **Dietterich, T. G.** (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation*, **10**(7), 1895–1923.

[8] **Doane, M. A.** (2003). The close-up: scale and detail in the cinema. *Differences: A Journal of Feminist Cultural Studies*, **14**(3), 89–111.

[9] **Ester, M.**, **H.-P. Kriegel**, **J. Sander**, and **X. Xu**, A density-based algorithm for discovering clusters in large spatial databases with noise. *In Kdd*, volume 96. 1996.

[10] **Eyben, F.**, **M. Wöllmer**, and **B. Schuller**, Opensmile: the munich versatile and fast open-source audio feature extractor. *In Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010.

[11] **Guha, T.**, **C.-W. Huang**, **N. Kumar**, **Y. Zhu**, and **S. S. Narayanan**, Gender representation in cinematic content: A multimodal approach. *In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, 2015.

[12] **Guha, T.**, **N. Kumar**, **S. S. Narayanan**, and **S. L. Smith**, Computationally deconstructing movie narratives: an informatics approach. *In Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015.

[13] **Jerrick, D.** (2013). The effectiveness of film trailers: Evidence from the college student market. *UW-L Journal of Undergraduate Research, XVI*, **13**.

[14] **Jia, Y.**, **E. Shelhamer**, **J. Donahue**, **S. Karayev**, **J. Long**, **R. Girshick**, **S. Guadarrama**, and **T. Darrell**, Caffe: Convolutional architecture for fast feature embedding. *In Proceedings of the ACM International Conference on Multimedia*. ACM, 2014.

[15] **Joshi, M.**, **D. Das**, **K. Gimpel**, and **N. 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*. Association for Computational Linguistics, 2010.

[16] **Kaya, N.** and **H. H. Epps** (2004). Relationship between color and emotion: A study of college students. *College student journal*, **38**(3), 396.

[17] **Kernan, L.**, *Coming attractions: Reading American movie trailers*. University of Texas Press, 2009.

[18] **LeCun, Y.**, **L. Bottou**, **Y. Bengio**, and **P. Haffner** (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, **86**(11), 2278–2324.

[19] **Levi, G.** and **T. Hassner**, Age and gender classification using convolutional neural networks. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2015.

[20] **Mestyán, M.**, **T. Yasseri**, and **J. Kertész** (2013). Early prediction of movie box office success based on wikipedia activity big data. *PloS one*, **8**(8), e71226.

[21] **Mishne, G.**, **N. S. Glance**, *et al.*, Predicting movie sales from blogger sentiment. *In AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*. 2006.

[22] **Ouyang, W.**, **P. Luo**, **X. Zeng**, **S. Qiu**, **Y. Tian**, **H. Li**, **S. Yang**, **Z. Wang**, **Y. Xiong**, **C. Qian**, *et al.* (2014). Deepid-net: multi-stage and deformable deep convolutional neural networks for object detection. *arXiv preprint arXiv:1409.3505*.

[23] **Panaligan, R.** and **A. Chen** (2013). Quantifying movie magic with google search. *Google Whitepaper-Industry Perspectives+ User Insights*.

[24] **Parkhi, O. M.**, **A. Vedaldi**, and **A. Zisserman** (). Deep face recognition.

[25] **Pedregosa, F.**, **G. Varoquaux**, **A. Gramfort**, **V. Michel**, **B. Thirion**, **O. Grisel**, **M. Blondel**, **P. Prettenhofer**, **R. Weiss**, **V. Dubourg**, **J. Vanderplas**, **A. Passos**, **D. Cournapeau**, **M. Brucher**, **M. Perrot**, and **E. Duchesnay** (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.

[26] **Reinstein, D. A.** and **C. M. Snyder** (2005). The influence of expert reviews on consumer demand for experience goods: A case study of movie critics*. *The journal of industrial economics*, **53**(1), 27–51.

[27] **Roos, D.** (2008). How movie marketing works. Available at `http://entertainment.howstuffworks.com/movie-marketing.htm`.

[28] **Sadikov, E.**, **A. Parameswaran**, and **P. Venetis** (2009). Blogs as predictors of movie success.

[29] **Schroff, F.**, **D. Kalenichenko**, and **J. Philbin**, Facenet: A unified embedding for face recognition and clustering. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

[30] **Schuller, B.**, **S. Steidl**, and **A. Batliner**, The interspeech 2009 emotion challenge. *In INTERSPEECH*, volume 2009. Citeseer, 2009.

[31] **Septak, D.** (2008). *Understanding the Effectiveness of Trailers, Teasers, and Television Spots in Marketing Movies*. Ph.D. thesis, Stern School of Business New York.

[32] **Sharda, R.** and **D. Delen** (2006). Predicting box-office success of motion pictures with neural networks. *Expert Systems with Applications*, **30**(2), 243–254.

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

[34] **Smith, S. L.**, **M. Choueiti**, and **K. Pieper** (2014). Gender bias without borders: An investigation of female characters in popular films across 11 countries. *USC Annenberg*, **5**.

[35] **Taigman, Y.**, **M. Yang**, **M. Ranzato**, and **L. Wolf**, Deepface: Closing the gap to human-level performance in face verification. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014.

[36] **Turner, B.** (2015). Why is a movie's first week box office so important? Available at `http://entertainment.howstuffworks.com/why-is-movie-first-week-box-office-important.htm`.

[37] **Wu, X.**, **R. He**, and **Z. Sun** (2015). A lightened cnn for deep face representation. *arXiv preprint arXiv:1511.02683*.

## LIST OF PAPERS BASED ON THESIS

1. **Adarsh Tadimari, Naveen Kumar, Tanaya Guha and Shrikanth S. Narayanan**. Opening big in box office? Trailer content can help. In IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), 2016.