Movie Subtitles Analysis - Understanding the Role of Emotional Dialogues in the Movie's success

Sri Harsha Pamidi (A59005361), Srinivas Rao Daru (A59003596)

1 INTRODUCTION

Movies and Television series have always played an important role in the entertainment industry. They include a variety of crafts such as cinematography, direction, screenplay, writing and so on, as well as a variety of genres. We attempt to understand some of the emotional driving factors in making a successful film using the script and the interactions between the characters as these are the most crucial elements of the movie. We also seek to comprehend some of the specific traits or properties that are required for success in each genre and understand the driving forces behind some of the most popular movies and movie series, as well as the recurring elements that have contributed to their success. We also aim to analyze the works of individual directors and cinematic universes in order to grasp their emotional preferences. We feel that their preferences for a specific emotion is one of the driving force behind their achievement. We also feel that most popular movies contain unique scenes that are the driving factors for their success and these are generally a strong concentration of a single emotion.

2 EMOTION FROM SUBTITLES

The emotional content of a dialogue is determined by both the words spoken and the semantic context of the scene. Subtitles give us the most exact interpretation of the script of a movie and we believe that by combining the words and the context in which they occur, we would be able to detect emotional content in groups of sentences. By context, we imply the tone of the conversation and it is pretty obvious from the subtitle text itself. So we collected subtitles(".srt") files from various openly available sources like opensubtitles, kaggle, then extracted the dialogues associated with the movie time, and performed sentiment analysis on them in order to comprehend the various emotions expressed throughout the film.

The data is then analyzed to determine the factors that contribute to a film's success as measured by its imdb rating. To observe the similarities and statistically significant factors, we used various techniques like correlation analysis and investigated various trends in the data across different films. Also, we try to analyse the subtitles of different movies and movie series to understand the flow of emotions throughout the movie.

3 UNDERSTANDING AND PRE-PROCESSING OF DATA

Emotion extraction from text is the most important aspect of the task because its accuracy affects how good the subsequent emotion analysis is. In the first part of the next section, we examine the various methodologies we used to calculate the emotion score, as well as their benefits and cons. Finally, we describe the algorithm we used to compute the emotion score.

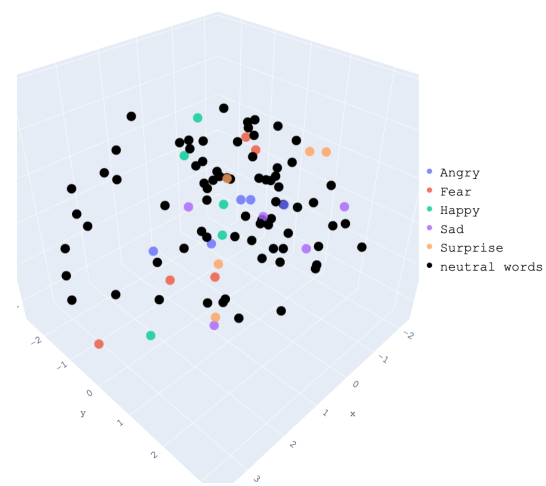


Figure 1: Visualizing Emotion Word Embedding with PCA and t-SNE

3.1 Pre-trained word embeddings

We used pre-trained GloVe word embeddings[2] for the emotion lexicons [3]. These word vectors are trained on corpus - Wikipedia 2014 + Gigaword 5 (6B tokens, 400K vocab, uncased, 100d vectors) and is made available under the Public Domain Dedication and License v1.0.

GloVe is an unsupervised learning algorithm for obtaining vector representations for words. Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resulting representations showcase interesting linear substructures of the word vector space.

We want to see whether we can establish a decision boundary between the word embeddings of different emotions, therefore we plotted the word embeddings and visualized whether the emotion words are clustering together. We illustrated the word embedding in **Figure 1** by employing common dimensionality reduction techniques such as PCA and t-SNE. To convert the words' vector representations in embedding space.

PCA is commonly used for dimensionality reduction by projecting each data point onto only the first few principal components to obtain lower-dimensional data while preserving as much of the data's variation as possible. t-distributed stochastic neighbor embedding (t-SNE) is a statistical method for visualizing high-dimensional data by giving each data point a location in a two or three-dimensional map.

We can observe in **Figure 1** that we were unable to discover any appropriate decision boundaries between the emotional words. So using these embed dings to create a vector for each word would not give any significant information about the emotion in the new dimensional space.

3.2 Word based comparisons

Many of the words generally have a sense of emotions and are in fact used to portray a specific one. So we have considered an approach in which we try to assign a specific emotion score to a scene based on the relative number of different emotion portraying words in them. One such technique using word based emotion detection has been discussed in [1]. The paper also discusses about different type of emotions in humans and also about some of the emotions which are a sum of multiple once. For emotion information the NRC lexicon from [2] has been used, which is a exhaustive list of words and related emotion collected by manually tagging and collected in open source. In this method first we have split the subtitles into a set of 3 minute scenes and then calculated the count of different emotions by calculating number of words that portray the emotion which is given in the NRC lexicon. We have observed that only less than 20% of the total words in the movie are present in the lexicon. The words which are not present are actually the words with no emotion, for example words like has been, character names, verbs and many others, and is not because of the lexicon words list. So the count itself cannot be scaled to find the emotion score in the scene. In the next section we discuss a package available in python which returns a list of emotion scores based on complete text which we believe is a better procedure in finding the related emotion.

3.3 Data Preprocessing

The first layer includes the removal of special characters, numbers, and excess spaces. Stop words were eliminated in the second layer. Furthermore, all letters in all words are changed to lower case and then to their base forms (stemming), and the subtitles are created into bins of 3 minute scene dialogues. To parse the SubRip (srt) files, we used the python subtitle-parsing module **pysrt**.

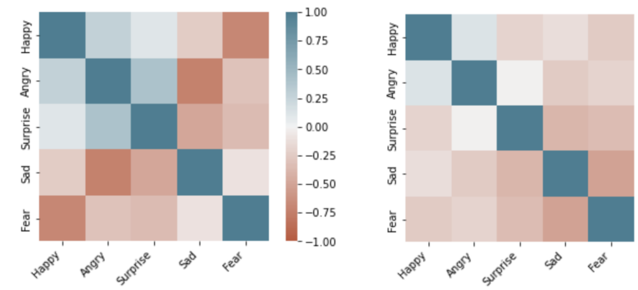
3.4 Emotion Score Calculation

To extract the emotions from the content, we utilized the text2emotion python package. It analyzes the dialogue in the movie scene and recognizes the emotions embedded in it, returning five separate emotion category confidence scores for Happy, Angry, Sad, Surprise, Fear and others.

Text2emotion cleans the data as explained in **Section 3.3** section and extracts the text suitable for emotion analysis by deleting unwanted textual content from dialogues and applying natural language processing algorithms to extract the well-pre-processed text from the text pre-processing.

It then detects the emotion from each word extracted from preprocessed text and counts it for further evaluation by finding the right words to explain the sentiment of the scene, examining each word's emotion category, and keeping track of the number of emotions associated with the words discovered.

Following that, emotion analysis is performed to determine the significant output for the dialogues in the bin. The end result will be a dictionary in which the keys represent emotion categories and the values reflect emotion scores. The higher the score of a certain



(a) (b)

Figure 2: (a)-(b) Correlation heatmap of different emotions for the movie Iron Man(2008) for total scene counts of n_scenes=10 for (a) and n_scenes=50 for (b)

emotion category, the more likely the discourse corresponds to that category.

3.5 Scene Time Frames and Emotions Used

A movie has numerous temporal frames in which the story centers around a different emotion in each one. Rather than assigning a single emotion value to the entire film, it is preferable to divide the film into sections and find the scores in each one. For this, we may either use a predetermined time gap between two scenes or divide the entire movie into a specified number of bins. The later is employed here since it is preferable to have vectors of the same size when attempting to detect a correlation between movies.

In addition, several of the films are approximately the same duration. A normal movie is roughly 100 minutes long, and we want a scene time of 2 minutes, which is a typically a scene length, thus **n_scenes** (the number of total scenes or time frames) was set to 50. We also attempted lower values, but as shown in **Figure 2(a),2(b)**, smaller values of n_scenes have a larger correlation (positive correlation) of distinct emotions than higher values of n_scenes. We anticipate that emotion values will be negatively linked rather than uncorrelated or positively correlated. Because, as previously said, we expect a scene to focus around a specific emotion, the prevalence of one emotion should reduce the occurrence of other emotions in that scene.

Humans have a plethora of emotions, and employing them all would be exhausting. We chose a few of them to offer a general picture of the setting, as well as those feelings that are unrelated to the situation. We can see from the **Figure 2(b)** that emotions are either uncorrelated or negatively correlated, which tells us that the emotions chosen would offer us a better grasp of the scene as it means that a scene revolves around only one of the given emotion.

3.6 Data Example

As shown in **Figure-3**, we extracted 50d vectors for each emotion (Happy, Angry, Surprise, Sad, Fear) throughout the movie. The x-axis represents the movie bins, and the y-axis represents the associated emotion ratings; notice that this is not a probability, because, as previously stated, there are a significant number of neutral moments in the movies. We can see that the movie has a surprise

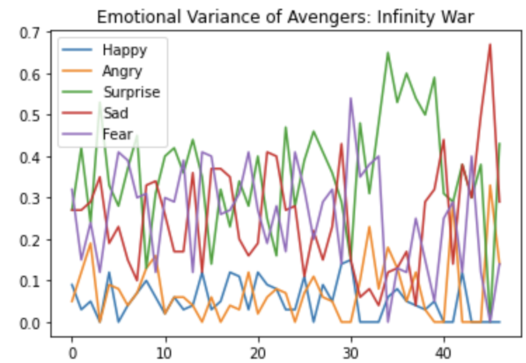
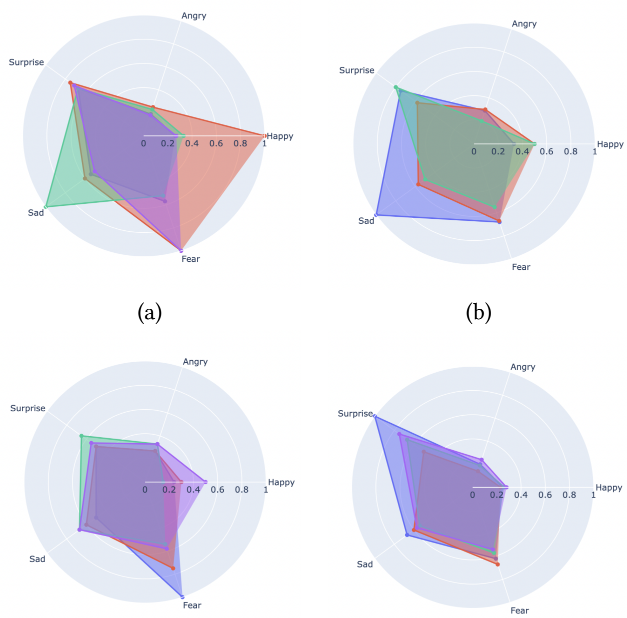


Figure 3: Emotional Variance over the time of movie Avengers: Infinity War (2018)



(a) (b)

(c) (d)

Figure 4: Plots of aggregated emotion vectors with various movies

emotion peak in the pre-climax and a sad emotion high in the climax. This corresponds to the action scene between Thanos and the Avengers, while the sorrowful climax corresponds to Thano's snap.

4 EXPLORATORY ANALYSIS

We analyzed various sets of films, such as sequels and films set in the same universe and written and directed by the same director, to see if there are any parallels or trends in the emotional experience to the films. In the **Figure 4** we have taken maximum of the 50d values in order to show the emotional bias in the genre as it corresponds to the most important scenes.

4.1 Comparison of films in a film series

Figure 4b depicts the aggregated emotional levels for the Iron Man films Iron Man (2008), Iron Man 2 (2010), and Iron Man 3 (2013). We can observe that all of the emotional scores are similar throughout all of the films, with a sorrow peak in the Iron Man 2008 film. Such common emotional mixture is not found in movies taken from multiple genres and which are unrelated like in **Figure 4(a)**.

4.2 Comparison of films set in the same universe (marvel)

Figure 4c shows that all of the Marvel Avengers films - The Avengers (2012), Avengers: Age of Ultron (2015), Avengers: Infinity War (2018), and Avengers: Endgame (2019) - are based on the same emotional scale, with the fear component ascending. Also, all of the emotional scores are consistent across all of the films, with the most fear peaking in the Avengers 2008 movie. This mixture is also similar to the Iron man movies coming from the same universe. The only difference being in the most preferred emotion which was fear in Avengers series whereas it was sad in Iron man movies. But within the series a specific emotion was preferred making them driving factors for the success.

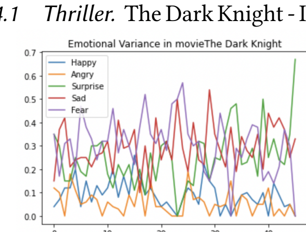
4.3 Comparison of films written and directed by the same director (Christopher Nolan)

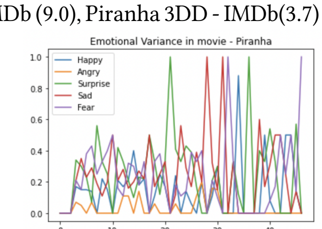
Figure 4d implies that most of Christopher Nolan's recent films are built on the same emotional scale, with more surprise components in the films. We compared the films Tenet (2020), Inception (2010), The Dark Knight (2008), and Interstellar (2014), and they are all nearly identical, with the largest surprise component in the tenet film.

4.4 Comparison of successful and unsuccessful films in the same genre

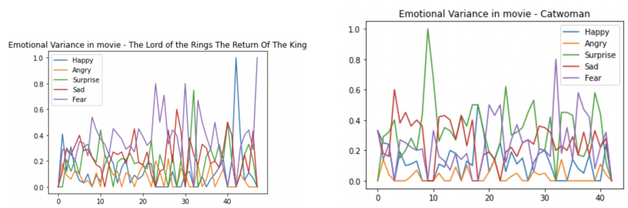
Figure 4(a), which compares the movies The Dark Knight (2008), The Lord of the Rings: The Return of the King (2003), Back to the Future (1985), and David Attenborough: A Life on Our Planet (2020), shows that the pattern of emotion scores is similar for movies in the same genre but varies depending on the genre. As a result, we decided to contrast and compare them within each prominent genre. We examined at the emotional scores of two movies from each genre, one with a high IMDb rating and one with a low rating.

4.4.1 Thriller. The Dark Knight - IMDb (9.0), Piranha 3DD - IMDb(3.7)

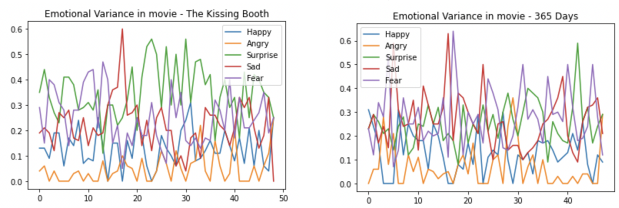




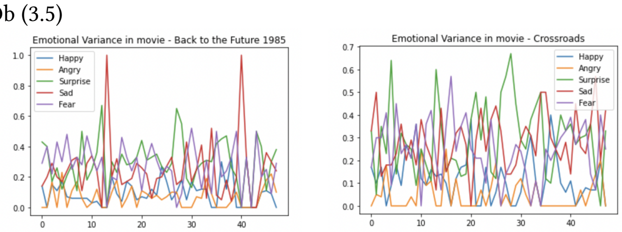
4.4.2 Action. The Lord of the Rings: The Return of the King - IMDb (8.9), Catwoman - IMDb (3.4)



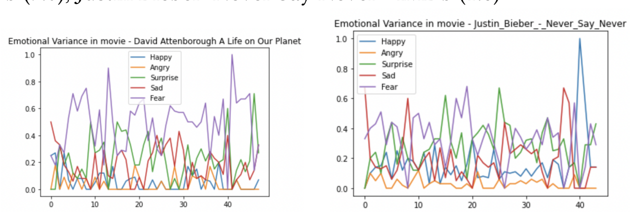
4.4.3 *Romantic.* 96 (II) - IMDb (8.5), 365 Days - IMDb (3.3)



4.4.4 *Comedy.* Back to the Future - IMDb (8.5), Crossroads (I) - IMDb (3.5)



4.4.5 *Documentary.* David Attenborough: A Life on Our Planet - IMDb (9.0), Justin Bieber: Never Say Never - IMDb (1.6)



4.5 Summary

Every movie has a high point of emotion. A comedy genre movie, for example, has a greater mean of happy vector than other emotions. This pattern can be found in any film where one or more emotions peak at a specific period in the narrative. This trend is seen in the figures of various genres depicted above. In addition, every film includes many peaks related to the most essential scenes. Peaks are less common in highly rated films and are typically seen in the middle and near the finish of the film. This indicates that directing the most concentrated sequences toward the conclusion would increase the success of a film. We can also presume that a small number of high-variation emotions would be preferable over a constant depiction of high emotions throughout the film.

5 CORRELATION BETWEEN EMOTIONS AND IMDB SCORES

The Pearson correlation coefficient is used to illustrate the relationships between emotions and user scores to movies on the IMDb database in the Correlation heatmap depicted in **Figure-5** and **Table-1**.

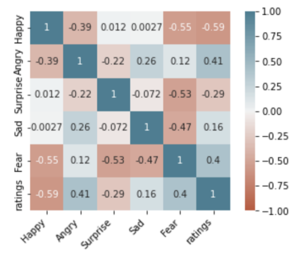


Figure 5: Correlation heatmap of aggregated emotions in relation to movie ratings and other emotions

Table 1: Pearson Correlation between Emotions and IMDb ratings

Emotion	Correlation
Happy	-0.59
Angry	0.41
Surprise	-0.29
Sad	0.16
Fear	0.40

As we expected, all the emotions have a clear connection with either high or low IMDb ratings. One of the most interesting relationships is the strong association between IMDb ratings and the movie emotions of rage and fear. Furthermore, we can see from the Pearson correlation results that positive sentiment - Happy and surprise emotions have a negative correlation with Users IMDb ratings. It contradicted our intuitive predictions because terms with positive sentiment, such as joy and anticipation, are assumed to correlate positively with people evaluations.

Based on these findings, we may conclude that an increase in any of the positive emotions of happiness or anticipation usually results in a decrease in other emotions and, resulting in the drop of user's liking for the movie and vice versa for negative emotions.

6 CONCLUSION

We discussed about several principles we may follow to make a movie more successful, such as having a director work on a specific movie emotion that he is skilled at or having dramatic changes in emotion, preferably in the climax sequences. It is better for movie series to continue with the same mix of emotions and to focus on certain feelings that lead to the success of the first movie in the series. It is evident from our work that we may analyze the success of a movie by examining the subtitles, and that the temporal change of emotions is vital in the success.

REFERENCES

[1] Amir Kazem Kayhani, Farid Meziane, and Raja Chiky. 2020. Movies Emotional Analysis Using Textual Contents. In *Natural Language Processing and Information Systems*, Elisabeth Métais, Farid Meziane, Helmut Horacek, and Philipp Cimiano (Eds.). Springer International Publishing, Cham, 205–212.

[2] Saif Mohammad and Peter D. Turney. 2013. Crowdsourcing a Word-Emotion Association Lexicon. *CoRR* abs/1308.6297 (2013). arXiv:1308.6297 http://arxiv.org/abs/1308.6297