



UNIVERSITÀ DEGLI STUDI DI MILANO

Emotion Detection in Harry Potter Movies

Text Mining and Sentiment Analysis Project Report

Serbu Elena Valentina

Student Number: 988351

Submitted: July, 2023

Contents

1	Introduction	1
1.1	Data Exploration	1
2	Emotion Detection Techniques Description	2
2.1	Approach 1	2
2.2	Approach 2	3
2.3	Approach 3	4
2.4	Approach 4	5
3	Social Network	7
4	Conclusions	9

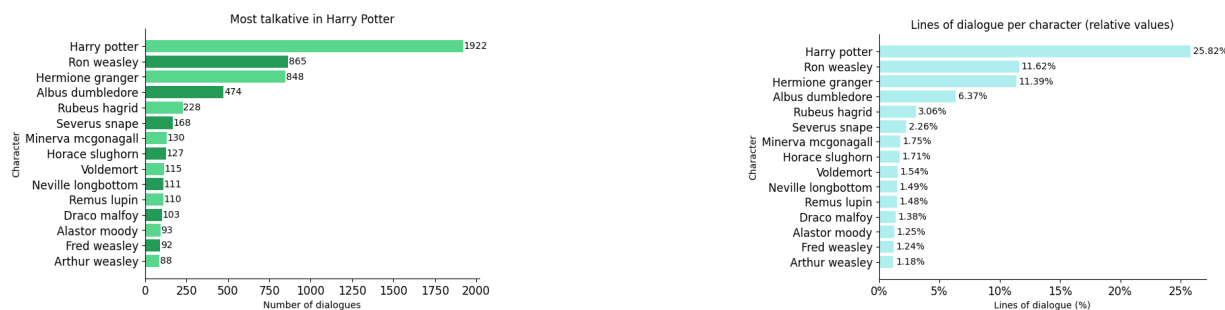
1 Introduction

The aim of the project is to creatively use emotion detection techniques to characterize the emotional and psychological profile of the characters and effectively represent their characteristics, their evolution in the story, and the interaction with other characters.

As a fervent fan of the Harry Potter series, I have always been intrigued by the profound emotions experienced by the characters throughout their journeys. This fascination, combined with my interest in data analysis and machine learning, led me to embark on a project centered on emotion detection using Harry Potter datasets. The characters in the series, such as Harry Potter, Hermione Granger, and Ron Weasley, undergo a range of emotional experiences, from fear and sadness to joy and love. By analyzing their dialogues and interactions, we can extract valuable patterns and characteristics that help us understand the emotional landscape of the wizarding world.

1.1 Data Exploration

Before starting with the actual Emotion Detection task, I will perform a brief Data Exploration in order to familiarize with the Dataset. The data contains 173 Characters, for a total of 8 movies. Since we will extract the emotions directly from the Sentences said by each character it is worth investigating which characters talk the most or have more lines throughout the movies.



As we can see from the plots above Harry, Ron and Hermione are both the characters with

more lines and that talk the most. It is interesting to notice that a character as important and intriguing as Dumbledore talks so little.

2 Emotion Detection Techniques Description

Emotion detection, also known as sentiment analysis, is the process of identifying and understanding human emotions expressed in text, speech, or other forms of communication. Emotion detection has applications in fields such as customer feedback analysis, social media monitoring, market research, mental health analysis, and human-computer interaction, among others. My Emotion Detection will contain:

- The overall distribution of emotions by character.
- The Character's Emotional progression throughout the movies.
- The character's Emotional progression throughout the Chapters of each movie.
- Building of a Custom predictive model for Binary Text Classification.
- Building of a Social Network between Movie's characters and referred characters in speech.

2.1 Approach 1

The aim of the first approach I decided to implement in order to perform Emotion Detection was to find an overall distribution of emotions by character. The characters I chose for this kind of Analysis are widely considered to be the most influential in the Movie's plot. So i joined all the sentences said by these influential characters, splitted them into singular words and applied a Sentiment Intensity Analyzer(SIA) to each word. A SIA is a rule-based sentiment analyzer in which the terms are labeled as per their semantic orientation

as either positive, negative or neutral. Once I got the sentiment associated to each word, I obtained the following plot:

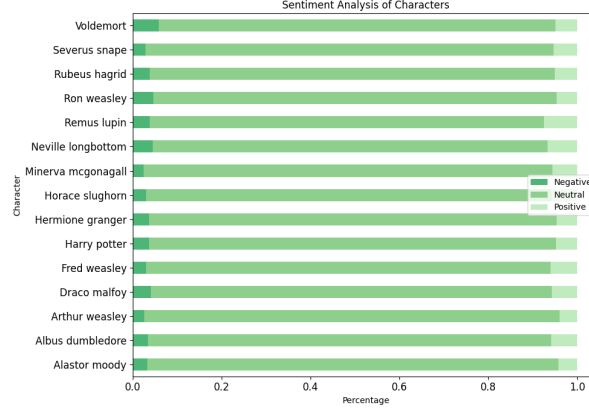


Figure 1: Emotion Distribution by words

As we can see from the plot above, when having a look at the words said by a character without a context, Voldemort is the most "evil" character. This kind of approach is a bit restrictive since it considers each word separately and it does not give much importance to the sentence as a whole. Though it is still interesting to see that characters widely known for being evil, also use words that are associated with negative feelings.

2.2 Approach 2

The goal of the second Approach I decided to implement in order to perform Emotion Detection on Harry Potter was to see the Emotional Journey of some of the main Characters throughout the Movies. The model classifies the text inputted into 5 emotions: ['love', 'joy', 'sadness', 'anger', 'fear']. After I applied the Classifier on each Sentence in the Dataset, I found the normalized count for the emotions for each character and Movie Part. The following is the Plot I obtained by selecting only Harry Potter and Voldemort. The plots below are very truthful in my opinion; Harry is a teenager that goes through seemingly impossible experiences and, as

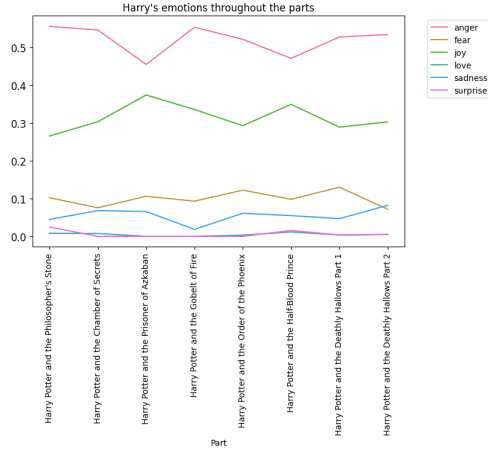


Figure 2: Harry's Emotional Journey

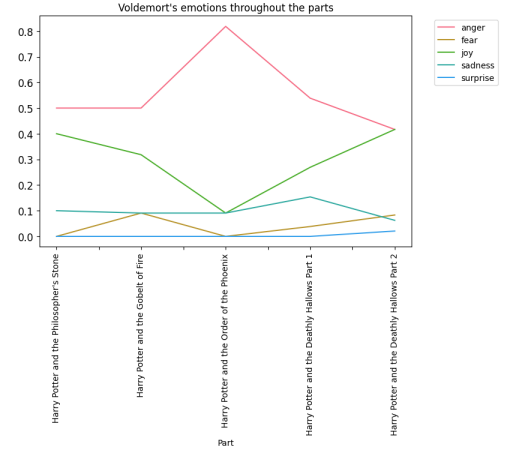


Figure 3: Voldemort's Emotional Journey

we can see from the plot, his main emotions are anger, joy or fear. Voldemort on the other hand is a cruel and gruesome character and in the plot we can see that he totally lacks 'Love' as an emotion. Also, in Harry Potter and the Order of the Phoenix, he undergoes a series of crucial changes that probably impacted on his joy.

2.3 Approach 3

The goal of the third approach was to see the Emotional journey of some of the main characters throughout the movies for each chapter. In order to perform this kind of Analysis I had to retrieve an index associated with each movie and chapter, so to have a final ordered Dataset. Then I decided to use the polarity score obtained by running the Sentiment Intensity Analyzer on the sentences. The polarity score (which ranges between 0,2) gives us the Intensity of the main emotion expressed by the sentence and not only the Emotion Label associated with it. In this way we can visualize the Emotional journey of the characters in each chapter along the Movies. The value of the Emotions depicted by the plot range between 0-2, zero being a very negative emotion and very positive emotion. As the Saga progresses we can see that Harry's Emotion shift from having very low points(meaning

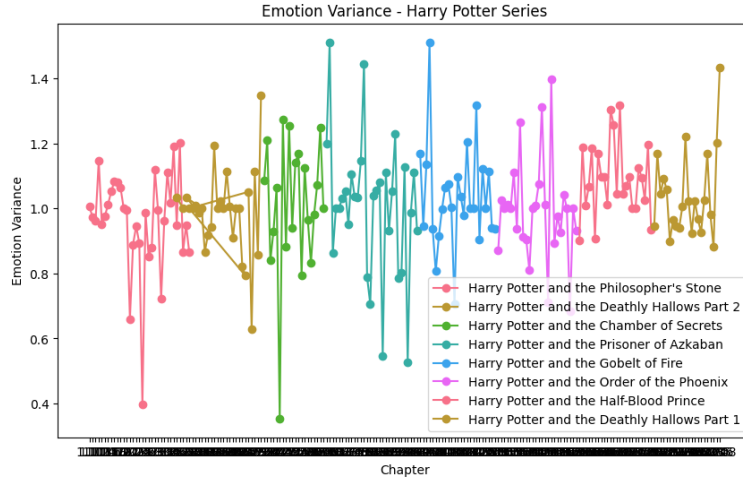


Figure 4: Harry’s Emotional Journey by Chapters and Movie parts

negative emotions) to having more positive emotions. This could be do to the fact that Harry grows and maturity can also be associated with a stricter range of emotions.

2.4 Approach 4

In this approach I tried to build a custom predictive Model for binary text classification(POSITIVE/NEGATIVE). A predictive model for text classification is a deep learning model that is trained to classify text documents into predefined categories or labels. It learns patterns and relationships within the text data to make predictions about the class or category to which a given document belongs. When such a predictive model is specifically used for sentiment analysis, it aims to determine the sentiment or emotional polarity expressed in a piece of text. The sentiment can be classified into categories such as positive, negative, neutral, or even more fine-grained emotions. In order to implement such a model I had to follow the following steps:

- Data gathering: Labeled existing Dataset using SIA algorithm, which means I associated to each sentence a sentiment. The dataset is then used for training and evaluating the predictive model.

- Text Preprocessing: The sentence Column has to be tokenized in order to proceed with the model build.
- Feature Extraction: The Preprocessed data which will be needed for model training needs to be extracted as numerical features.
- Model training: I trained a sequential Neural Network, which i built using the keras library.
- Model Evaluation

The following is the summary of the Model i built for text classification: I trained the

Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 192, 10)	61230
flatten (Flatten)	(None, 1920)	0
dense (Dense)	(None, 64)	122944
dense_1 (Dense)	(None, 1)	65

=====
Total params: 184,239
Trainable params: 184,239
Non-trainable params: 0
=====

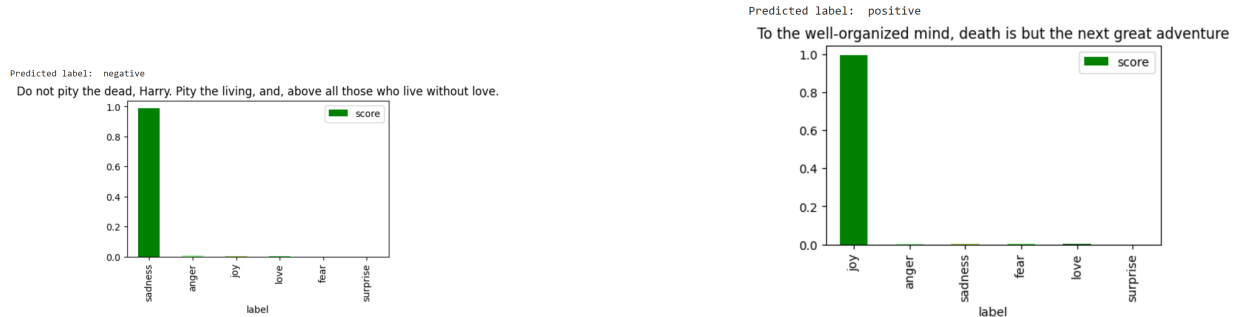
Figure 5: Model summary

model for 10 epochs, using a batch size equal to 256. The results I obtained after evaluating the Model on a test set are the following:

Test loss: 0.3881 - Test accuracy: 0.8489

The Model I implemented is not ideal since we have a high validation loss compared to the training loss. A high validation loss indicates that the model is not able to generalize well to new, unseen data. In other words, the model is performing poorly in terms of minimizing the error or discrepancy between its predictions and the true labels on the validation dataset. Despite these results I decided to proceed by applying my Model to

new unseen sentences, in order to see if it was able to understand the general direction of the phrase's feelings. In order to double check my model's prediction, I also plotted the



emotions extracted from the sentence using the DistilBert Model. As we can see from the plots the predictive model I implemented predicts correctly the sentence's emotion since in the first phrase the main emotion is sadness(NEGATIVE) and in the second one the main emotion is joy(POSITIVE)

3 Social Network

The necessary requirement for building a Social Network from a Dataset of Movie Dialogues is that there must be some sort of interaction or relationship between the considered characters. The initial Dataset I was working with just contained the Character's Name and the Sentence they said; I added a column with the name, when present, of another Character that appears in the sentence. For example if Professor Dumbledore said:" Harry how are you?" then the referred Character would be "Harry potter". Using the networkx Library I built a directed graph where the nodes are the Characters that referred to someone, the edges between them are the referral relationship with its direction and the emotion of the sentence is an additional attribute of the edge.

```
# Add an edge between the characters involved in the dialogue
G.add_edge(character1, character2, emotion=emotion)
```

The resulting graph has 104 nodes and 229 edges, with only 47 mutual edges. Characters like Harry Potter, Dumbledore, Hermione Granger and Ronald Winsley have very high degree, in-degree and out-degree, confirming the fact that they are also very influential figures inside the Harry Potter Movies; on the other hand Voldemort has a very high in-degree but low out-degree, which basically means that he is the main topic of most conversations in the series. The node representing Harry Potter qualifies as a hub when considering degree, betweenness centrality and page rank. This result seems reasonable since the whole series revolves around his character. From the plot above we can see that

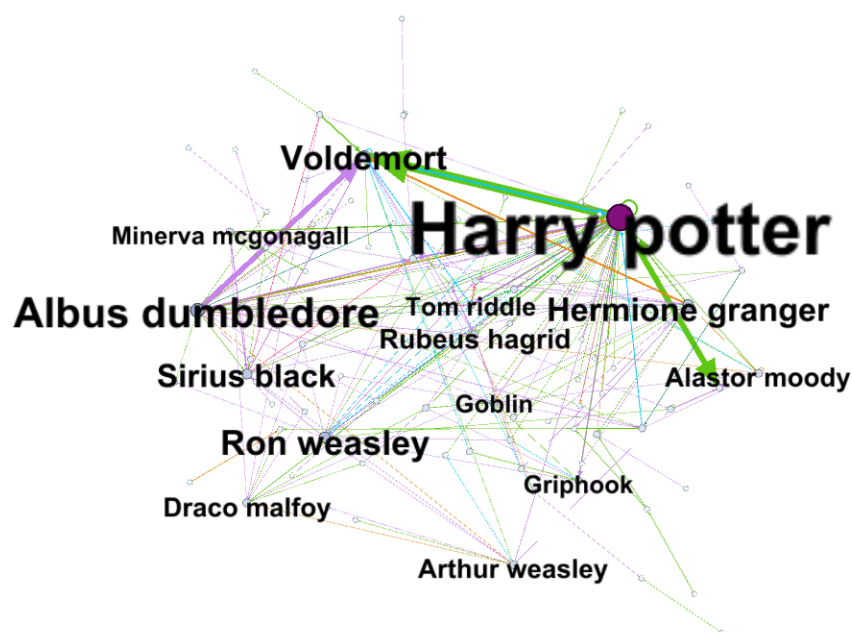


Figure 6: Social Network

both Dumbledore and Harry Potter, when referring to Voldemort are moved by Anger or Fear.

4 Conclusions

In this project, I focused on emotion detection in the Harry Potter movies to gain insights into the characters' emotions and how they align with the main plot assumptions. By analyzing the characters' dialogues and sentiments expressed throughout the movies, I was able to draw several conclusions that confirmed the initial knowledge I have on the movies plot. Firstly, I found out that there is a wide range of emotions throughout the Harry Potter movies. Characters exhibited emotions such as joy, anger, fear, and more, reflecting the complexities of their experiences and the challenges they faced. Furthermore, my analysis revealed consistent patterns of emotions in alignment with the main plot assumptions. For instance, during critical moments of conflict or danger, characters often displayed heightened levels of fear and anxiety, emphasizing the high stakes and suspenseful nature of the plot. Finally the social Network confirmed the emotional dynamics i have seen in the movies. This project lays the foundation for further research on emotional storytelling and its impact on audience engagement and emotional connection with fictional characters.