How do you feel, my dear

Gabriele Fusar Bassini[1]

Università degli studi di Milano
gabriele.fusarbassini@studenti.unimi.it

Abstract. The aim of this paper is to exploit NLP tools and emotion detection in order to analyze the structure of movie scripts. We argue that dialogues reflect the emotion the characters, hence their changes through time. We can therefore proceed in understand how those are related to particular narrative styles. We develope a model with Support Vector Machine on a dataset classified on a slightly modified Ekman's six basic emotions. Then, we apply the model on scraped scripts. The analysis is enriched by the case study of "Annie Hall" (1977) by Woody Allen, a love story with no happy ending, testified by the emotional behaviour of its protagonist.

Keywords: Emotion detection · Movies · Dialogues.

1 Introduction

Over time, several methods have been proposed for finding recurring patterns in narratives. These are forms of stories that have their roots in the oldest tales, of which authors such as Syd Field have identified the structural elements, such as the division into acts and the moments of climax and relaxation. More recent research has instead highlighted the importance of the protagonist's point of view, as in the case of Christopher Vogler's "The Writer's Journey". In his book, the American author selects some passages and the ways in which they occur, as well as the emotional response of the protagonist. Nowadays, the progress of emotion detection allows us to look for the same patterns in a considerable amount of examples through what transpires from the different dimensions of the emotional identity of the characters starting from their dialogues.

1.1 Emotion Detection

Natural Language Processing (NLP) is a branch of Artificial Intelligence focused on treating and analysing human language. Its core intent is to make machines capable of understanding and producing human language. NLP implies several types of resources, such as images and audios, but the availability of big datasets thanks to the Web 2.0 make textual data a great opportunity in the field. [1] An important aspect of communication is emotion and it can be expressed in different forms. Text-Based Emotion Detection (TBED) has been used by researchers to automatically detect affect, identifying the feelings and sentiments expressed in a text. It utilizes methods of computational linguistics, text analysis, machine learning, and NLP to detect emotions from various data sources. [2] Emotions are usually represented either as categorical variables, such as Ekman's six basic emotions [3] or in a continuous space. The latter refers to metrics such as valence (psitivity/negativity) and arousal (the intensity of the sentiment). In this case, we use a hybrid approach: we choose to rely on a model similar to Ekman's, but we exploit the scores of each variable (i.e. each emotion) as a proxy for intensity to use them as continuous variables.

2 Research question and methodology

The main goal of this paper is to infer informations on the plot and narrative techniques of movies starting from the emotions leaked from the protagonists' words as time series. The project is divided into two main parts: first, we build a model starting from a dataset of almost 20000 phrases annotated with a sentiment between anger, fear, joy, love, sadness and surprise [4]. In the second stage, we import a single movie (as this is a preliminary work to a more cospicous one). We manipulate it until we select only a polished number of quotes of its main character and proceed in scouting the emotional changes through time. That is, we apply the model derived before to the quotes and plot them divided into the six emotions. This is an explorative work mainly interested in the qualitative results of the metodology.

3 Experimental results

3.1 NLP Dataset

Different scholars have put their effort into creating a series of dataset consisting in texts and annotated labels. Those classifications vary in form and number. A notable example is Emotion Detection from Text [5], a corpus of 40000 tweets with relative tag, divided into 9 different possibilities. Altough interesting in number, the choice for this project fell on a more cautious dataset. The first step while working on this kind of corpus is to clean the tweets from usernames, links and other characteristics component of the written language that we won't find in the spoken utterances. The process is straightforward, but introducing a dictionary that transforms slangs into spoken words is not enough to capture the complexity of the Internet jargon. However, the main issue here is the complexity of the labels: highly unbalanced, their number may not get along with the simplification needed in order to investigate movies. For this reasons, and for a lack of better performance in comparison with the NLP dataset, it is better to choose a dataset that better fits our needs. The emotion dataset consists of almost vairous 20000 utterances with no context but the emotion transmitted. It uses a classification similar to the famous Ekman's, except from having a label for love and no label for disgust. The dataset is imbalanced in some classes (surprise especially), but overall quite balanced between positive and negative emotions, with most quotes classified as joyful or sad. The operations performed on the dataset are quite standard: we clean them from unwanted characters, transform abbreviations into common words with the use of a dictionary [6], perform lemmatization and delete stopwords. Lemmatization comes in handy with the dimensions of this dataset: it would in all likelihood not contain many forms of the same word that we may find later in the dialogues. Avoiding stopwords is instead a good choice when we need for more defined results, since stopwords are common with every emotion and would tend to neutralize the sentiment of an utterance.

3.2 Movie Corpus

Despite the fact that we focus on a single movie in this report, the movie dataset prepared for this report is an extensive one. It collects almost 3000 movie scripts of various genres, also tagged by year and additional metadata. Cornell Movie-Dialogs Corpus [7] is a tempting alternative: it contains contains a large metadata-rich collection of fictional conversations extracted from raw movie scripts, with 220,579 conversational exchanges between 10,292 pairs of movie characters from 617 movies. Praised for its variety and used in different project, including a notorious PyTorch Tuturial for a Chatbot, the Cornell Dataset doesn't however contain the entire spoken quotes of its characters, making it diffult to track the emotional arc during time. That's why, in order to have a more in depth analysis, we need create a dataset of quotes from scratch, starting from scraping the original movie scripts. Luckily, there are different online resources to get them. The dataset chosen is primarily built on the online resource IMSDb. The limitations of a dataset built through scraping are quickly apparent: some scripts are unusable due to incorrect download. Others differ greatly in their use of conventions. The most efficient implementation, after analyzing different formatting methods, is a three-step process. The first one includes the majority of scripts, i. e. the ones providing the most formal layout and some cases with a white line splitting each couple of written ones. The two subsequent approaches, deployed only if the first one fails, try to catch dialogues written in different forms. Almost 2700 of the scripts pass this part, since many return no useful information.

Since the main goal in this phase is to extract the lines spoken by the main characters in the films, we must start with some assumptions. In particular, the main theoretical assumption, confirmed by preliminary empirical analysis, is that the main character is the one with the most number of dialogues. Although this is not a certainty, it is reasonable to assume that with a dataset of several thousand films, there will be few cases that contradict this hypothesis. We also assume that the scripts are formally correct and therefore always indicate the main character with the same name. Even in this case, we cannot take care of some exceptions such as changes of identity, but these are extremely limited cases. An additional assumption is that there are no misleading indications in the scripts that make non-existent characters appear, outside of examples such as CUT. TO, INT., EXT. and other common stage notes.

3.3 SVM Model

The model created with the NLP dataset has a quite remarkable internal performance. After vectorizing the now lists of words for every phrase with TF-IDF (term frequency-inverse document frequency), we are able to feed a supervised machine learning algorithm (in this case linear Support Vector Machine is working best). The choice of Scikit Learn SVM is not only fast and reliable, but also easily serves our second goal. We can in fact retrieve the single scores for every label instead of having only the one outputted by the algorithm. Thanks to this trick, we transform the categorical labels into continuous ones without loss of meaning (as we will assess later, for example negative emotions will have similar patterns). To evaluate the performance a 5-fold cross validation is sufficient to assess the model good behaviour. The training accuracy is 0.974 ± 0.002, while the test accuracy is 0.856 ± 0.005. On the overall dataset, while the training accuracy slightly decreases, the test betters. This enlights two facts: the model is working good, as it looses bias with a larger datset. Moreover, a richer corpus may even improve the performance. The table below assesses a good precision in the larger classes, while the model suffers more when the dataset contains lesser examples of a certain emotion. Neverthless, the performance is still acceptable. Love, with 74% of correct labelling, has the lowest accuracy.

Table 1. Precision of the model on the test data for each sentiment

sentiment	precision	recall	f1-score	support
anger	0.84	0.88	0.86	283
fear	0.83	0.79	0.81	231
joy	0.88	0.91	0.89	660
love	0.74	0.64	0.68	159
sadness	0.90	0.91	0.90	585
surprise	0.85	0.81	0.83	79
accuracy			0.86	1997
macro avg	0.84	0.82	0.83	1997
weighted avg	0.86	0.86	0.86	1997

3.4 Analysing the movies

Once the protagonist quotes are selected and cleaned just like the dataset phrases, we can pass them to our model and, once processed, obtain for each emotion a time series that explains its trend over time. Clearly, we cannot expect but a chaotic behaviour: a single sentence can easily betray the sentiment of a phase, which is instead considered by looking at the larger picture. For this reason, in translating these data into exploratory graphs, it is necessary to use a moving average, which in this case preserves the memory of the forty previous sentences. Although this is a limitation of the model, which cannot in fact analyze a significant part of the film (its beginning), such a moving average allows us not to be at the mercy of sudden changes due to classification errors or small sentences whose importance is limited in the overall emotional arc. The specific movie considered in this analysis is "Annie Hall" (1977) by Woody Allen. The sentimental comedy is one of the most notorius example of the peculiar style of the American author. In the movie, Allen interprets Alvy Singer, a New York comedian that looks at the past to understand his relationship with Annie Hall, with whom he broke up a year before. Alvy's sentiment pass through a lot of emotions such as sorrow, joy, love and jealousy. Paired with a classical yet witty plot, it makes a good candidate to explore the behaviour of the model created before. As Fig. 1 depicts, love is the most fleeting emotion: its pattern, apart from the rise before the end of the first act, is pointing downwards, leaving space for negative emotions. Another interesting plot is fear: rising from the beginning of the movie, as anger does, it has three main movements, roughly coincidental with a three acts movie structure. Joy is opposite to sadness: growing in the beginning and mainly high during the developement of the movie, it falls down in the end when Alvy desperately tries to win Annie back, while anger takes over. Lesser appears to say about surprise: poorly represented in the original dataset, its behaviour may be affected by a worse evaluation than the others. Its significance is also of minor importance during time: surprise is a fleeting emotion, we hardly see it as a persistent sentiment. That may justify its peculiar plot, presenting more spikes than the others.

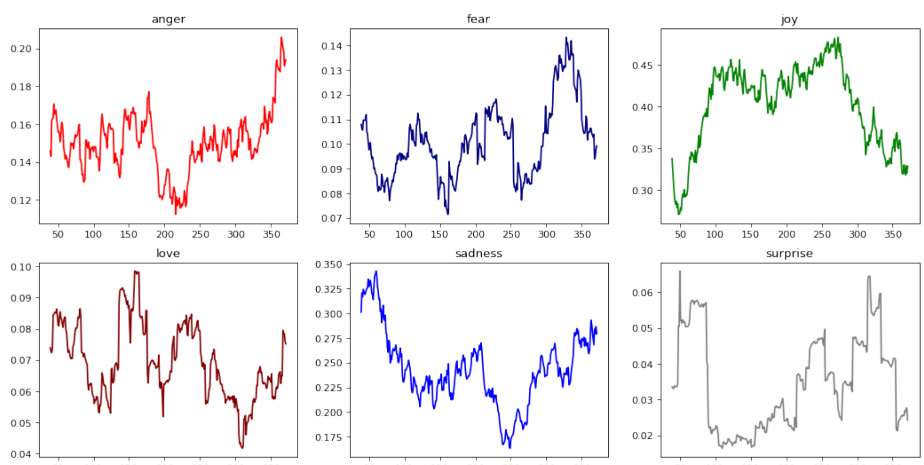

Fig. 1. Emotions arcs in "Annie Hall": anger, fear, joy, love, sadness, surprise

3.5 Concluding remarks

The model based on NLP dataset confirms to perform well, with an accuracy of 86% in the test set. It is able to classify speeches with no pre specified length into six different sentiments. Using SVM, we are able to transform those categorical label into continuous ones; this allows us to scout their behaviour through subsequents utterances treating them as time series. Once we have scraped every quote from a character in a movie in chronological order, we are therefore able to understand the shift between emotional states through the plot events. The work on "Annie Hall" is reflected in the script of the movie and we can find several qualitative informations on the emotional response of the protagonist Alvy to the events of the movie. This paper contributes to find an automatic method to select the alleged main character of a movie and her/his lines thanks to a specific framework to treat screenplays. It also has some promising insight on how movie structures and hero's journeys relate to their expressed emotions. The exclusion of external notes from the dialogue itself is a strong assumption, since it elides important information from the model. Neverthless, we obtained the expected result, that induce us to think that dialogues preserve enought information to conduct an analysis. A further work would focus on expanding the model on IMSDb's entire dataset, looking for recurrent patterns through movies. One would expect to see specific emotional arc in certain structures, such as the fall of positive feelings in a tragedy, or an abrupt inversion of feelings in the notorious last act resolutions. Since this exploratory analysis witnesses the importance of the duality between positive and negative emotions, it may be useful to move it to an approach that exploits valence instead of those six labels. This choice would enhance the possibility to compare movies on the basis on single time series.

References

1. Acheampong, Francisca Adoma and Nunoo-Mensah, Henry and Chen, Wenyu: Transformer Models for Text-Based Emotion Detection: A Review of BERT-Based Approaches. Kluwer Academic Publishers (2021)
2. Zad, Samira and Heidari, Maryam and Jones, James and Uzuner, Ozlem: Emotion Detection of Textual Data: An Interdisciplinary Survey. IEEE World AI IoT Congress (AIIoT) (2021)
3. Ekman, Paul: An Argument for Basic Emotions. Cognition and Emotion (1992)
4. Emotions dataset for NLP https://www.kaggle.com/datasets/praveengovi/emotions-dataset-for-nlp. Last accessed 27 Jan 2023
5. Emotion Detection from Text, https://www.kaggle.com/datasets/pashupatigupta/emotion-detection-from-text. Last accessed 27 Jan 2023
6. Up-to-date list of Slangs for Text Preprocessing, https://www.kaggle.com/code/nmaguette/up-to-date-list-of-slangs-for-text-preprocessing/notebook. Last accessed 27 Jan 2023
7. Cornell Movie-Dialogs Corpus, https://www.cs.cornell.edu/ cristian/Cornell_Movie-Dialogs_Corpus.html. Last accessed 27 Jan 2023