Emotion Exploration using Multiple DL Classifiers

Università degli Studi di Milano (January 26, 2024)

Shihab Hamati (985941) and Alfredo Funicello (02675A)

Abstract

This report describes the work done in the context of the project for the Text Mining and Sentiment Analysis course by Prof. Alfio Ferrara at the University of Milan in 2023. The dialogue of all 8 seasons of a popular children show My Little Pony are used as the dataset. This dataset does not contain human-labelled emotions. However, thanks to modern widely available tools such as ChatGPT, we leveraged it using Prompt Engineering to create the emotion labels for us. We also used a Hugging Face emotion detection model as a second source of label generation. A third and more involved method of Fine Tuning a Distrilbert pre-trained LLM was also used. This third method allows for much more flexibility over the previous two, and this is shown by further tuning it for character detection (which can not be done by the off-the-shelf first 2 models). Finally, multiple analyses are made about the timeline of emotions in the storyline, for the characters, and an overall emotional profile.

Keywords: Emotion Detection, LLM, Large Language Model, Fine Tuning, Distilbert, ChatGPT

1 Introduction

This project aims (1) to explore modern machine learning tools in the context of text-based emotion detection, and (2) to utilize data science analyses to understand better how characters exhibit their emotions in a children TV show. The TV show is My Little Pony, a very popular children show, bringing in profits over USD 1 billion annually in 2014 and 2015. The dataset used is composed by 8 seasons worth of dialogue.

The dataset rows contain the dialogue line, the character speaking, the episode, and the writer of the episode; however, not the emotion of the utterance. Even a few years ago, it would have required costly and slow human labor to review over 25k sentences and label the emotions. This is how popular datasets like the Friends show was labelled. Employing machine learning, in this project, emotional labelling of the dialogues was completely automatized, consequently allowing us to analyze the emotions and character profiles of this body of work. In the context of the project we apply 3 emotional labelling methods:

- using an off-the-shelf HuggingFace emotion detection transformer
- Zero-Shot ChatGPT with Prompt Engineering

- Task-aware Fine Tuned DistilBERT LLM (through the Ktrain library)

The second part of the report covers the analyses and conclusions on the data gathered using the ChatGPT emotional labels. We explore:

- the overall emotional profile of the main characters
- the emotional timeline across seasons and episodes
- possible patterns in plots emotion distribution
- writers impact on emotion distribution in plots

2 Data Cleaning

The dataset was initially composed of 36859 dialogues from the 8 seasons of the show. The purpose of the following project is to analyze the emotional depiction of characters in the show using their dialogues. Some characters have less than 200 lines in the data, to reduce noise some of the data from these lower appearing characters has been dropped. Figure 1 displays the graph used to guide the data trimming decision; 0.68% of the initial data was kept after the trimming, resulting in a dataset composed by only the most appearing characters of the show of 25419 dialogues.

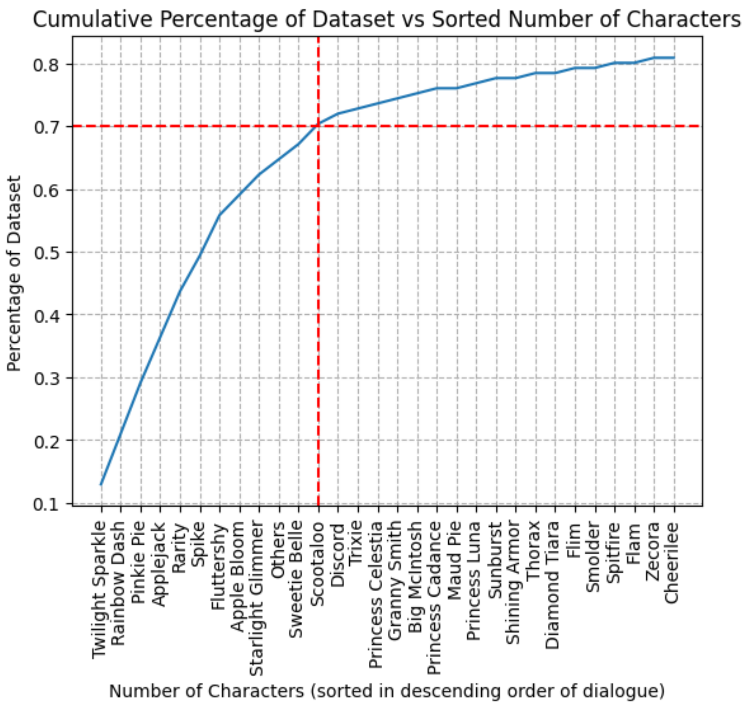

Fig. 1 Comulative distribution of dialogue across characters

3 Data Labelling

3.1 Using an Off-the-Shelf Emotional Detection Transformer

One convenient and fast method for emotion detection is to use a pretrained model from HuggingFace. There are multiple emotion detection models. We opted to use the transformer from j-hartmann (https://huggingface.co/j-hartmann/emotion-english-distilroberta-base). This is because it uses Ekman's 6 basic emotions (anger, disgust, fear, joy, neutral, sadness, surprise) plus neutral. This we felt is a balanced number of classes for our dataset size (from the lowest end of binary sentiment classes of positive and negative plus neutral, to more elaborate emotion classification theories with dozens of emotions).

This transformer was pretrained by it's creator on a balanced set of labelled datasets containing aroung 20k data points and achieving a 66% accuracy. It is quite popular with around 1.7 million downloads. This transformer, as well as the many others available on HuggingFace, are fine tuned versions of much larger and much more sophisticated models trained on bigger datasets. In this case, the transformer fine tuned a DistilRoBERTa, which is a distilled version of RoBERTa. RoBERTa is a large language model based on BERT and trained on a large corpus of English data in a self-supervised way (by randomly masking 15% of words in a sentence and learning to predict them).

3.2 Zero-Shot LLM Emotion Labelling (through prompt engineering)

Advances in pretraining on unlabelled data have brought up the potential of better zero-shot or few-shot learning using generalized LLM models for data augmentation. This shows promising performance on zero-shot NLP semantic extraction and labelling tasks using prompt based methods.

Given the vast support of research bodies on the matter and the handy availability of LLMs such as OpenAI models, through their APIs paid program, zero-shot prompt-based emotion labelling using the same Ekman's paradigm emotion labels has been considered an useful exploration in the context of this work.

The openai and the langchain python libraries have been employed for handling the communication with the OpenAI APIs. By using the ChatPromptTemplate langchain interface, which simulates a chat interaction, the gpt−3.5−turbo−1106 OpenAI model has been iteratively interrogated with a prompt composed by a system message used to deliver the task guidelines and an human message used to deliver the dialog entry from the data.

The guidelines message used was

```
You classify the emotions of this sentence into one of the following
["anger","disgust","fear", "joy", "neutral", "sadness", "surprise"]. You must
answer with a single word from the prior list. If you are not sure, return
neutral.|
```

followed by the dialogue entry to be labelled. The answered label was then added to the data.

In order to understand the cost of the labelling procedure the tiktoken python library was employed to estimate. OpenAI offers for its APIs a token-based pricing, with a cost of $ 0.001

every 1000 tokens for its gpt−3.5−turbo model. The tiktoken library uses the gpt−3.5−turbo tokenizer to estimate the amount of tokens in a body. Running the estimation over the data and adding the overhead of tokens created by the task guidelines message, which is repeated for every dialog, the estimate of the cost of the operation over 25120 dialog entries is shown in tab 1.

Table 1 Cost estimation of emotion labelling using OpenAI APIs

	Price	Percentage of total
Dialogue	$ 0.45	32.1 %
Guidelines	$ 0.94	67.9 %
Total	$ 1.39	

As the tab shows most of the cost was because of the overhead added by the guidelines, a methodology that could be further investigated for possible optimization by exploiting the token memory window of the model. Given OpenAI rate limits of 10000 requests per day the labelling procedure spanned 3 days.

3.3 Task-aware Fine-Tuned DistilBERT Emotion Labelling

An alternative more robust yet more involved method is to fine tune a pretrained large language model neural network such as BERT or DistilBERT to our specific task. In this case, the ChatGPT obtained emotion labels were used as the target column for a Text Classification task. The ktrain library was used as it provides a lightweight wrapper for Tensorflow Keras and helps rapidly build, train, and deploy neural networks.

We opted to fine tune a DistilBERT model because it is faster with lesser resource requirements as well as the lower chance of overfitting due to it being a smaller model compared to BERT and due to our small dataset (for NLP purposes).

The first step was to clean the data, for example by omitting non-letter characters. This helps constrain the dataset to more standard words that a pre-trained model would have been exposed to (and hence has hopefully learned some useful features about). Another step was to drop rows with emotions outside the 6+1 set mentioned previously. A small percentage of ChatGPT responses were outside the prompt-indicated bound (∼1%). Also, a maximum length parameter of 100 was used, and this is reasonable since as can be seen from the figure 2 almost the entire dataset falls into this constraint.

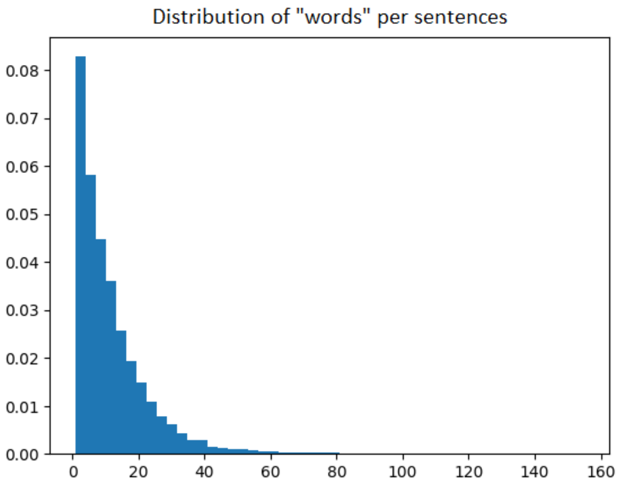

Fig. 2 Sentences length

When training a neural network, it is always vital to identify the ideal learning rate at which to do so. For this purpose, the ktrain function lr_find () was used over 2 epochs. From the graph 3, optimal learning rate suggestions vary from 8.32×10^{-7} to 5.56×10^{-5}. These correspond approximately to the longest valley, minimum gradient, and minimum loss, which have been shown to provide ideal choices to pass to the fitting function (used as an upper limit to the learning rate).

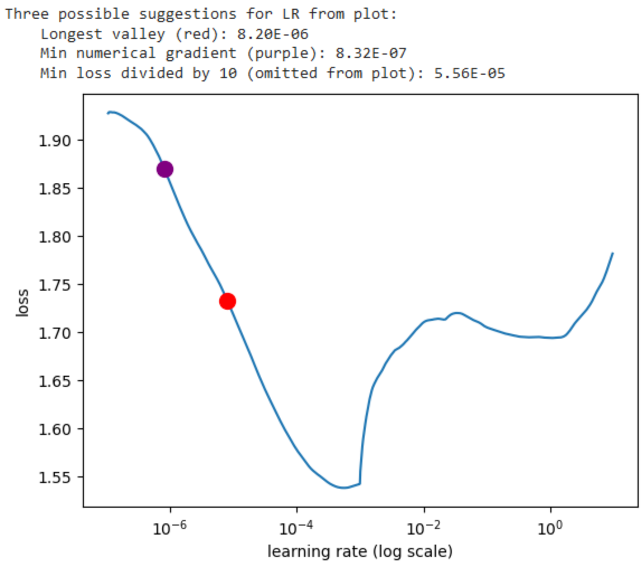

Fig. 3 Optimal Learning Rate

Since finetuning uses a pretrained model not a lot of time and resources are needed to adjust it to a new task. The training takes approximately 4 minutes per epoch. After training it for 5 epochs, the model gradually gains in training accuracy up to 91%. However, the model suffers from over-fitting, as these accuracy gains are not passed onto the validation set, as seen in the figure 4.

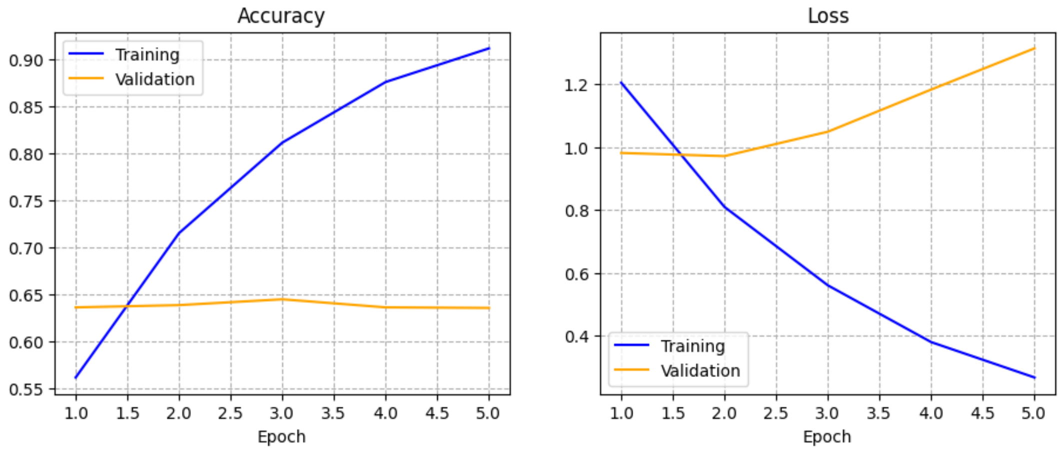

Fig. 4 Training Accuracy & Loss vs Epochs

Nonetheless, a validation accuracy around 64% is almost twice the largest class in the dataset (of about 39%). This indicates the learning power of this approach. Figure 5 and 6 are 2 outputs of the first couple of epochs when the learning rate is large and not optimal vs when it is optimal.

```
[33] eps = 5
     lr_opt = 1e-3 # for emotions_chatgpt

history = learner.autofit(lr=lr_opt, epochs=eps, checkpoint_folder="checkpoints")

begin training using triangular learning rate policy with max lr of 0.001...
Epoch 1/5
1256/1256 [==============================] - 258s 196ms/step - loss: 1.6660 - accuracy: 0.3883 - val_loss: 1.6325 - val_accuracy: 0.3903
Epoch 2/5
1256/1256 [==============================] - 242s 193ms/step - loss: 1.6525 - accuracy: 0.3900 - val_loss: 1.6311 - val_accuracy: 0.3903
Epoch 3/5
 767/1256 [=================>............] - ETA: 1:25 - loss: 1.6559 - accuracy: 0.3889
```

Fig. 5 Early epcohs without optimizing LR

```
[40] eps = 5
     lr_opt = 5.56e-5 # for emotions_chatgpt

history = learner.autofit(lr=lr_opt, epochs=eps, checkpoint_folder="checkpoints")

begin training using triangular learning rate policy with max lr of 5.56e-05...
Epoch 1/5
1256/1256 [==============================] - 260s 197ms/step - loss: 1.2051 - accuracy: 0.5618 - val_loss: 0.9814 - val_accuracy: 0.6361
Epoch 2/5
1256/1256 [==============================] - 242s 193ms/step - loss: 0.8094 - accuracy: 0.7149 - val_loss: 0.9714 - val_accuracy: 0.6385
Epoch 3/5
 362/1256 [=======>......................] - ETA: 2:35 - loss: 0.5188 - accuracy: 0.8299
```

Fig. 6 Early epochs with optimizing LR

In further steps, one might use a smaller model, increase the dataset size, or add counter measures to reduce this over-fitting problem. It is worth noting, that the popular HuggingFace transformer used in section 3.1 achieved a similar validation accuracy of 66% trained on a similar amount of 20k data points (albeit on a reportedly balanced set of datasets with a more varied context).

3.4 Task-aware Fine-Tuned DistilBERT Character Labelling

To illustrate one of the advantages of finetuning over ChatGPT and HuggingFace Emotion Detection Transformer approaches, a DistilBERT based model was trained to detect which character (pony) is speaking based on the dialogue.

After 5 epochs, the training accuracy is 65% while the validation accuracy plateaus at 37%. Although this might not seem much, it is worth remembering that the largest class when it comes to characters in the trimmed dataset is that of Twilight Sparkle at 18.6%. This exhibits the ability of finetuning to learn new classification tasks that can not be performed using off-the-shelf transformers or even prompt engineering (in theory it is possible, but it has to be fed so much data which might exceed practical API usage limits at the current time). There might also be a potential to improve the accuracy by tackling the overfitting in similar ways as suggested in section 3.3. Figure 7 the plots for optimal learning rate search and the training metrics.

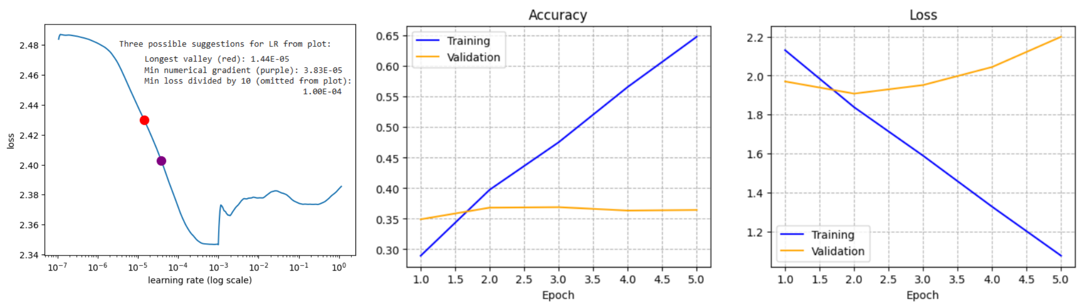

Fig. 7 Plots for Character Fine Tuning

4 Analyses of the emotionally augmented dataset

4.1 Characters' Emotional Profiles

Season 1 of the show is where characters get introduced and are most well characterized in order to get the viewer accustomed to the various different personalities, reason for which this analysis is focused on it.

To investigate if the emotional labels found by the model are able to depict accurately the personality of each of the characters, the distribution of emotions for each of them has been calculated and then normalized on a 0-1 scale using the max values found over all characters. How well this methodology describes the character will be explored by comparing the radar charts visualizations of the data and the My Little Pony Wikia pages of each of the characters. The most interesting finds are outlined below.

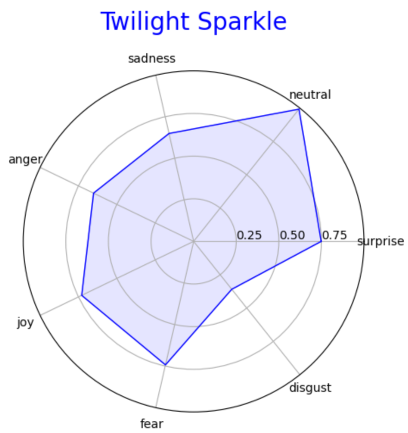

Fig. 8 Twilight's emotional radar chart over season 1

Twilight Sparkle is the main central character of the show, the plot of the episodes are usually recounted through her point of view of the events. Given her centrality in the show she is the most well balanced character. Her Wikia features entries such as *She tries to be rational in unfamiliar situations*, *Twilight tends to be skeptical of unproven claims* and *However, Twilight can lose her cool under stress*, just as it can be seen from the visualization from the data in figure 8, shes a balanced halfway point between other characters.

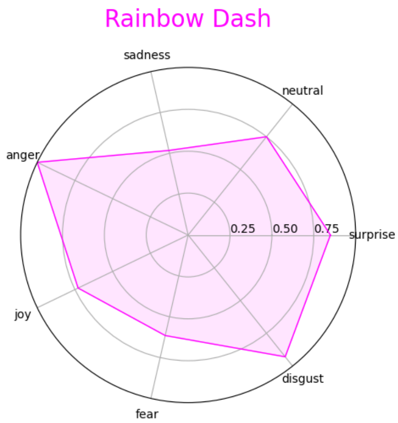

Fig. 9 Rainbow Dash's emotional radar chart over season 1

Rainbow Dash is described as conceited, often boasting, very competitive, sometimes mischievous. It seems that this traits lead her to be the character expressing the most anger, as shown in figure 9.

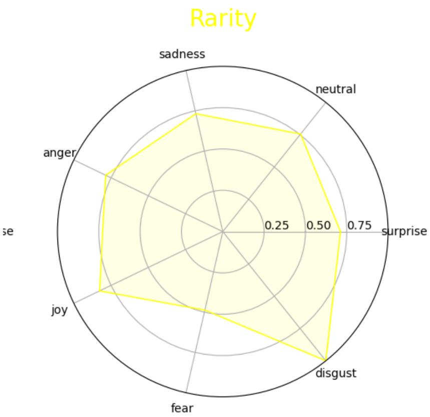

Fig. 10 Rarity's emotional radar chart over season 1

Rarity's Wikia entry describes her as *Rarity's mannerisms are similar to those of Scarlett O'Hara, the histrionic anti-heroine of Gone With the Wind. Many of Rarity's lines are rephrased from lines Scarlett says in the 1939 movie. Her vocabulary is formal, and she is prone to use complex words and more sophisticated, refined phrasing than her friends. As a fashionista, she often uses French-based terms in her vernacular. She speaks with a cultivated trans-Atlantic dialect, and shares some mannerisms with similarly accented Hollywood actresses, such as Katharine Hepburn.* Figure 10 show how her sophisticated and refined nature reflects in her emotion distribution, she's the character expressing the most disgust; probably because of her high standards and her dramatic personality.

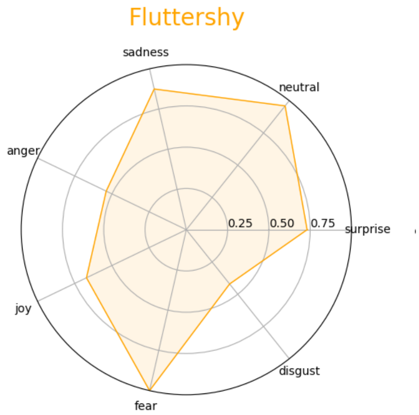

Fig. 11 Fluttershy's emotional radar chart over season 1

One of Wikia's entry for Fluttershy is *When Fluttershy is first introduced in the series, she barely manages to tell Twilight Sparkle her own name on account of her timidity*. Figure 11 displays her distribution, she's the most fearful and is often sad which aligns with her Wikia.

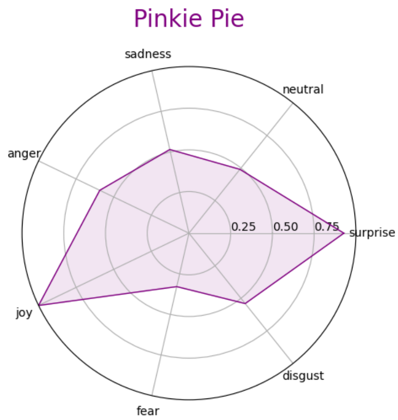

Fig. 12 Pinkie Pie's emotional radar chart over season 1

Wikia's entry for Pinkie Pie depicts her as *Pinkie is hyperactive, excitable, quirky, and outgoing, often speaking and acting in non sequiturs*; this entry also aligns very well with the character emotion distribution found in figure 12. She is the character expressing the most joy and displays a high amount of suprised dialogue.

This analysis provided an insight into how the writers of the show use the emotional tone conveyed in the dialogues by the characters to depict their distinct personalities. Each character's emotional reaction to plot events is not only a way to drive the narrative but aligns closely with their established personality traits and role within the series. The detailed portrayal of characters' emotional personalities is likely a deliberate choice to resonate with the show's diverse audience. Each character's unique emotional characteristics allows different viewers to connect with various aspects of their personalities, seeing parts of themselves reflected in the characters' experiences and reactions. This connection enriches the viewing experience, making the characters more relatable and the storytelling more impactful.

4.2 Emotional patterns throughout the seasons

Macro level analysis does not find any seasonal macro pattern but micro episode level analysis explains plots of episodes well.

To analyze the emotional evolution of characters over the seasons, we constructed a data structure which stores the percentage of emotional labels found in each episode's dialogue for each character. This methodical approach allows us to quantify and track the emotion expressed throughout the seasons.

```
{
    {
        ’episode_name1’:
            {
                ’character_name1’:
                    {
                        ’anger’: 2.5,
                        ’disgust’: 0.0,
                        ’fear’: 5.0,
                        ’joy’: 40.0,
                        ’neutral’: 30.0,
                        ’sadness’: 15.0,
                        ’surprise’: 7.5
                    },
                ...
            },
        ’season’: 1
    }
    ...
}
```

Analyzing the emotional labeling of dialogues over the seasons reveals a predominance of joy in the characters' utterances, a finding that aligns well with the show being aimed at young kids and the central theme of friendship and harmony the show has. As it can be seen in figure 13, a graph of emotions for the character Spike over season 1 of the show, the most prevalent emotion found is joy.

The analysis showed no relevant pattern of emotional evolution could be found in any of the characters when comparing between seasons. No signs of a trend might signify that the composition of episodes in a season doesn't generally follow any predefined plot scheme.

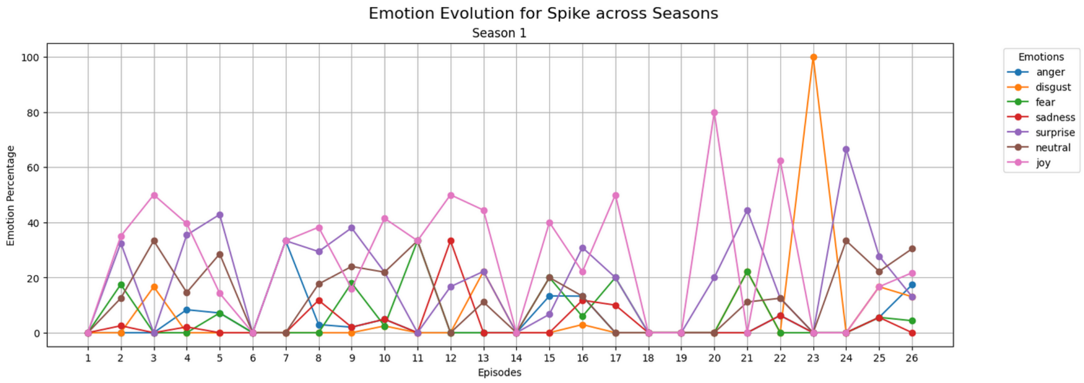

Fig. 13 Spike's emotion evolution across season 1

Instead of uncovering any specific patterns, this analysis approach interestingly highlights which episodes are more emotionally charged for characters. By only relying on labelled emotions over the dialogues of the episode it becomes evident which episodes feature heightened emotional experiences for the characters.

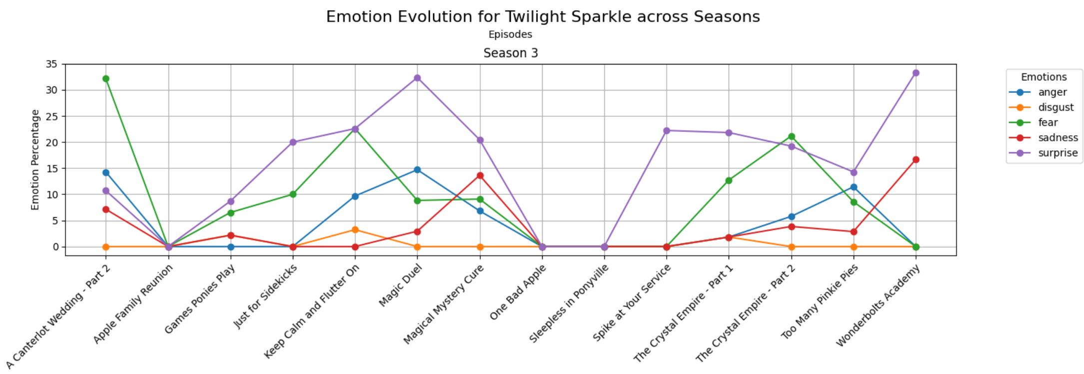

Fig. 14 Twilight Sparkle's emotion distribution for season 3 (joy and neutral are hidden due to their prevalence in every episode)

Peaks of *neutral* and *joy* in dialogues don't seem to highlight anything of relevance, given that neutral dialogues are probably just common interactions and joyful interactions seem to be general basic atmosphere of the show.

A clear example of this is on display in figure 14. The graph shows a notable peak in the emotion of *surprise* for Twilight Sparkle in episode "Magic Duel" of Season 3. This spike suggests that something occurred in this particular episode that caused astonishment or amazement from the character, which could be tied to a specific event or interaction within the storyline; the accompanying feelings of *anger*, *fear* and *sadness* suggest that Twilight might have been in a position of shock or dismay instead of positive surprise, and that the episode features a very negative setting. To fully understand the context of this emotional peak, the episode's content has to be inspected to see what might have triggered such a response.

Magic Duel. Season 3 *Seeking revenge the powerful Trixie returns to Ponyville, defeating Twilight in a magic duel. She exiles Twilight from Ponyville, who must figure out a way to best Trixie if she is to return home.*

The summary of the episode, indicating that the episode hosts an antagonist, in conjunction with the data gathered suggest that losing against Trixie and being exiled could have been the significant trigger event for Twilight's negative emotions. Data from other characters during the same episode highlights how even in negative episodes the language tends to be *joyful*, but given the negative events unravelling in the script most of the cast expresses out of the average levels of *fear* (in green) and *anger* (blue).

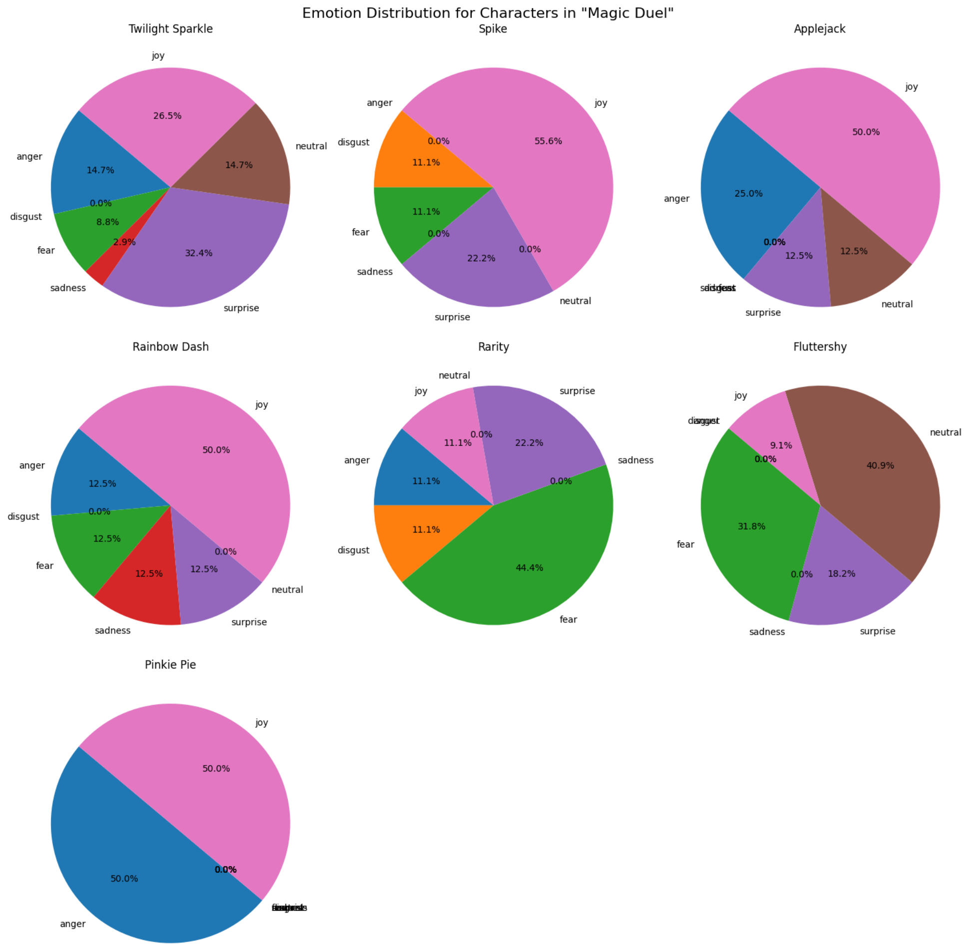

Fig. 15 Breakdown of emotion distributions for all characters during the episode "Magic Duel"

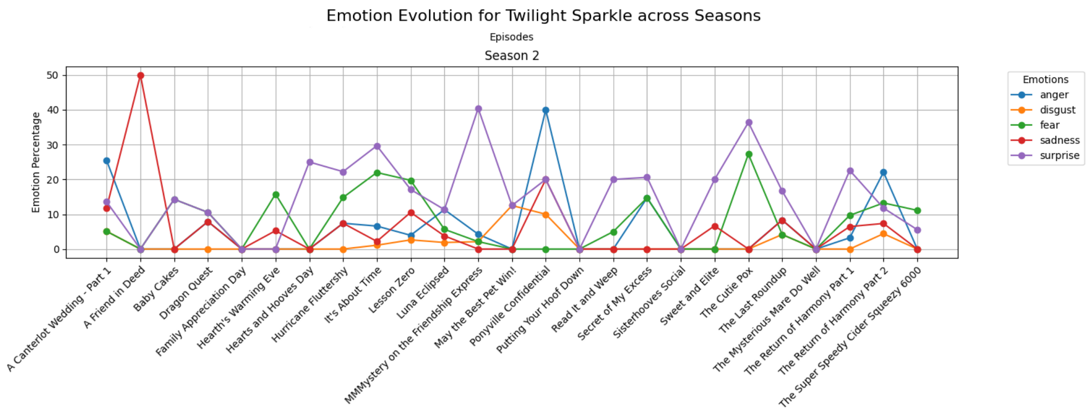

Fig. 16 Twilight Sparkle's emotion distribution for season 2 (joy and neutral are hidden due to prevalence in every episode)

The "A Canterlot wedding - Part 2" episode from season 2 displays another out of average peak in figure 16 for negative emotions.

A Canterlot wedding. Season 2 *Twilight saves her brother and all of the ponykind by freeing the real Cadence and defeating Queen Chrysalis, a changeling that had assumed Cadence's appearance in an attempt to take over Equestria.*

The summary confirms the presence of another struggle for Twilight Sparkle against an antagonist; by looking at the temporal distribution of emotion labelling for utterances over time in the episode in figure 17 the data also seems to imply that the episode ends on a good note aligning with the summary.

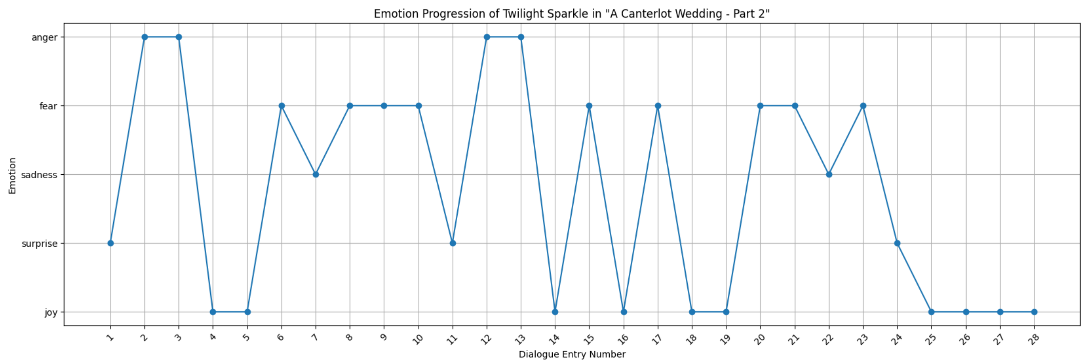

Fig. 17 Emotion labelled dialogue overtime by Twilight Sparkle in the episode "A Canterlot Wedding - Part 2". Emotions are ordered in descending order by negativeness.

Similar conclusions can arise from the analysis of episode "Ponyville confidential" from season 2; from figure 16 a big spike in *anger* along with *surprise*, *sadness* and *disgust* suggest that the character experienced particularly unpleasant circumstances which raised negative emotions. Looking at data from other characters shows that the general setting of the episode is once again negative and significative amounts of anger (in blue), sadness (in red) and disgust (in orange) are expressed.

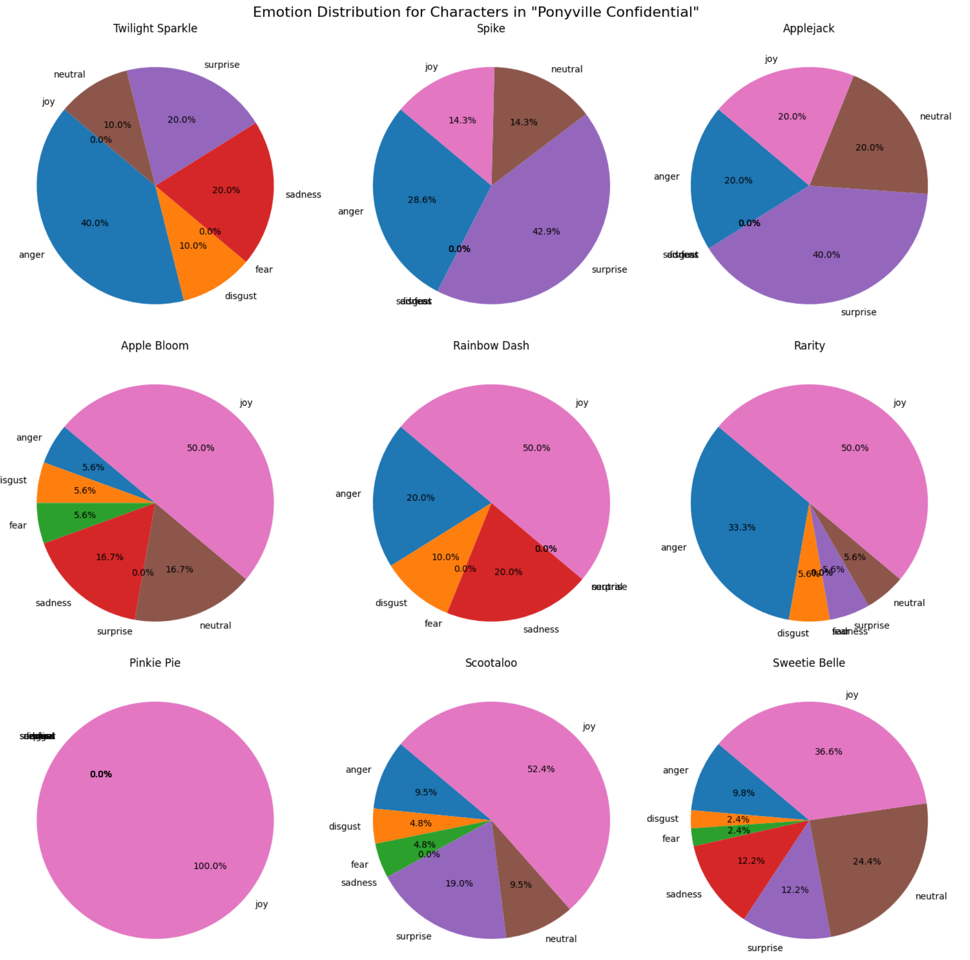

Fig. 18 Breakdown of emotion distributions for all characters during the episode "Ponyville confidential"

Ponyville confidential. Season 2 *The Cutie Mark Crusaders start a gossip column under the name Gabby Gums, but find it might not be worth the pain the stories cause other ponies.*

The summary confirms the hypothesis raised by the data; the main cast is being affected by unwanted gossip coming from a group of other characters.

The episode "Yakity-Sax" is an example where the micro episode-level analysis seems to be able to paint a very accurate picture of the plot of the story. By taking a closer look at the plot of the episode some of the discernible patterns in figure 19 can be explained.

Yakity-Sax. Season 8 *Pinkie Pie has a new hobby that she absolutely loves - playing the Zenithrash; when her friends discourage her from playing due to her lack of skill, it causes a series of events leading to Pinkie Pie possibly leaving Ponyville forever.*

The summary of the episode suggests that Pinkie Pie is the character central to the story.

{...} *Unfortunately, Pinkie's yovidaphone playing causes constant disruptions for her friends' daily activities. Fluttershy fears the sound is coming from a creature in trouble or pain. It distracts Rarity's sewing at Carousel Boutique, wakes up Fluttershy's animal friends from their sleep, interrupts Rainbow Dash's Wonderbolts flight show, and explodes all of the Apple family's harvested apples at Sweet Apple Acres into applesauce* {...}

In the top initial part of the graph the characters seem to express negativity, probably in relation to Pinkie Pie's hobby.

{...} *Despite having little talent playing, Pinkie greatly enjoys playing it, and her friends support her new hobby, believing she will improve* {...}

Wile expressing negativity most of the characters of the show try to continue approaching the issue in a positive manner, which explains the cluster of dialogues labelled as *joy* in the first part of the episode.

{...} *ponies eventually tell Pinkie that she's terrible at playing* {...} *Pinkie appears visibly shaken by this news* {...}

The negativity eventually gets to Pinkie Pie, which has multiple *sadness* labelled lines as it can be seen by the cluster towards the end of the episode.

{...} *Twilight and the others realize they were wrong to make Pinkie stop doing something she enjoys so much, and they encourage her to continue playing the yovidaphone as long as it makes her happy* {...}

The episode ends with a positive note, in fact the last cluster of dialogues is labelled as *joy*, it's also interesting to see, on the same time segment, the cluster of Pinkie Pie dialogues labelled as *surprise*, which suggests that the character gets surprised by the newfound support showed by her friends in response to her sadness.

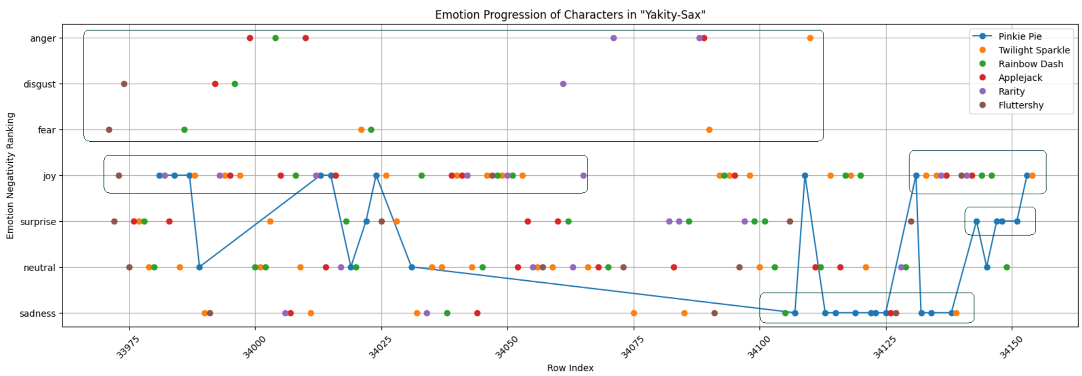

Fig. 19 Breakdown of emotion distributions for characters during the episode "Yakity-Sax"

This analytical approach for character emotion distributions over episodes seems to be a useful tool to detect shifts in the show's narrative, particularly when episodes diverge from the usual cheerful atmosphere to explore more challenging situations. By inspecting these emotional changes over the distribution, it becomes clear when the storyline is introducing conflicts or obstacles that the characters must confront, providing depth and complexity to the narrative and allowing for character growth and development.

4.3 Patterns of Emotion distribution over plots of episodes

The previous analysis suggests that some episodes are structured with a clear sequence of events that shape the story's progression. These events and character interactions drive the emotional shifts within an episode, creating multiple emotional segments which follow each other; this flow of emotions serves to advance the story in a way that is engaging and meaningful.

A hypothesis risen from this observation is that there might be a pattern in the way emotions are distributed across the timeline of the episode. This hypothesis is similar to https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-016-0093-1. The chosen methodology to explore the given hypothesis in the data is by subdividing episodes in 3 segments (start, middle and end) of equal size and calculating emotion distributions over them. Episodes where the title contained 'part' were excluded from the data assuming that the plot is spread across multiple episodes. Figure 20 is an example of the outlined methodology applied to the episode 'Yakity-sax' analyzed previously in section 4.2.

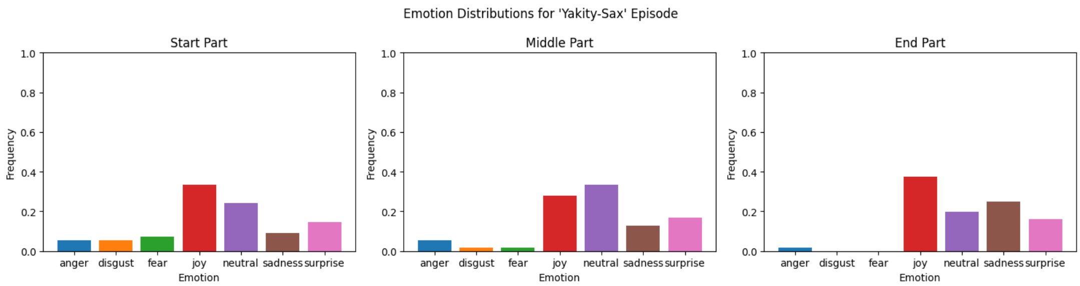

Fig. 20 Breakdown of emotion distributions for characters during the episode "Yakity-Sax" in different segments of the plot.

After review of the resulting data, given the high variance of distributions across the segments found, a higher abstraction method was needed. Binning the emotions in negatives (anger, disgust, fear) and positives (joy) was preferred under the assumption that narrative plot segments can be generally correctly simplified to negative atmosphere segments and positive atmosphere ones by looking at the emotions expressed through them. In table 2 the average value over all episodes of positive and negative percentages of emotions expressed in each segment. The show seems to start in most of the episodes on an overwhelmingly positive note. The middle part seems to be more balanced and the end part is the most emotionally charged segment usually.

Table 2 Negative and positive avg percentage of emotions over all episodes

Segment	Negative %	Positive %
Start	0.176 ± 0.006	0.444 ± 0.0178
Middle	0.240 ± 0.009	0.369 ± 0.013
End	0.276 ± 0.010	0.375 ± 0.011

Furthermore in order to understand if episodes could be clustered in groups having a similar negative-positive emotion distribution over all three narrative segments the kmeans algorithm was employed over the neg-pos percentage for each segment as feature for the episode. Using the elbow method on the WCSS plot in figure 21 3 clusters were identified.

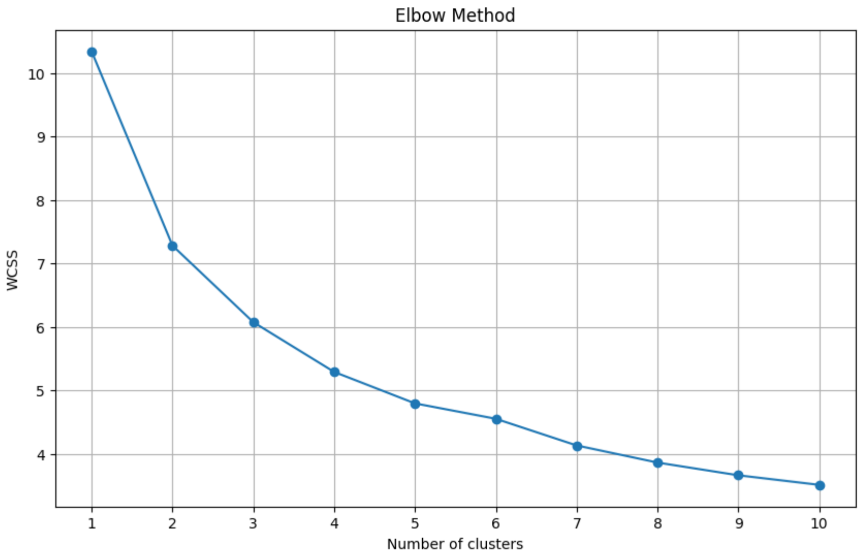

Fig. 21 Elbow method over the clustering of episodes using neg-pos segments.

The centroids in figure 22 describe the 3 clusters.

Cluster 3 is overwhelmingly positive episodes, the closest episode to the centroid is 'On Your Marks'.

On your Marks. Season 6 *Now that they've finally received their cutie marks, the Cutie Mark Crusaders struggle with the question of what's next; the friends do not all agree on how to embrace their destinies.* Receiving a 'Cutie Mark' is a sign that the pony character or skills have grown and is considered a symbol of 'purpose' for the character; receiving one is a very positive experience.

Cluster 2 is episodes where there's an equal balance between positive and negative from the start, the closest episode to the centroid is 'Scare Master'.

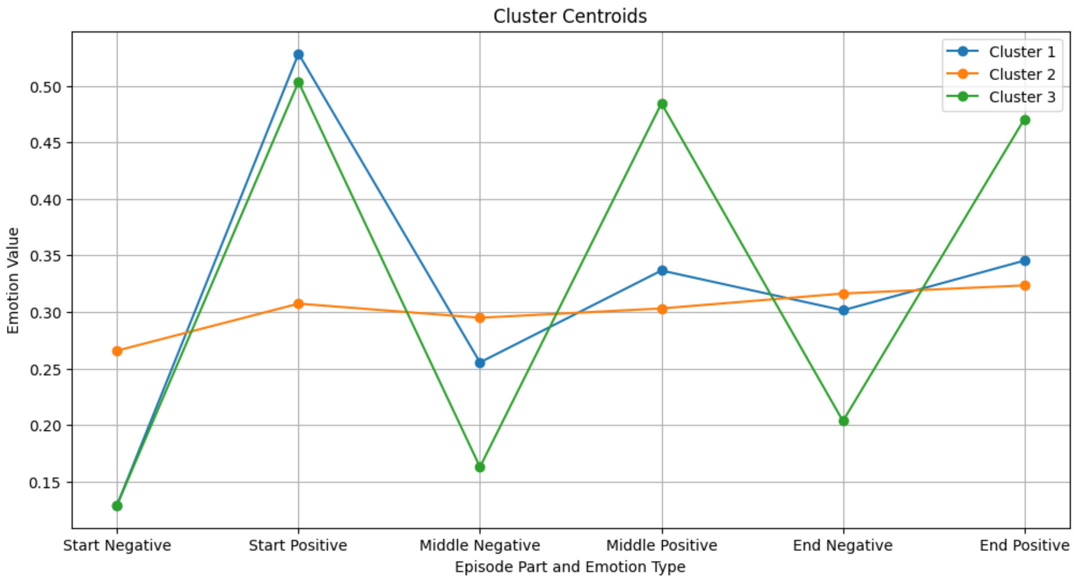

Fig. 22 Emotion value for centroids

Scare Master. Season 5 *Fluttershy is preparing to stay inside on Nightmare Night, but is forced to go outside when she discovers Angel has no food.* Fluttershy is a timid and fearful character, as evidenced in section 4.1, getting out for her during Nightmare Night (Halloween) seems to have been an all around negative experience introducing negative emotion dialogues right from the start.

Cluster 1 is episodes that start very positively, as most of them do, and then during the developing of the plot an event happens that introduces some negativity and reduces the initial positive atmosphere; the closest episode to the centroid is 'My Little Pony The Movie'.

My Little Pony The Movie *After their homeland is destroyed by Tempest Shadow, Twilight Sparkle and her friends embark on a journey to find the queen of hippos.* The movie seems to follow the very classic 'everything was very fine then came the bad guy' plot trope.

The PCA over the episodes paints an apparently well defined picture of the clusters. On the contrary the 0.20 Silhouette score scored by the clustering might suggest that the result might not be very statistically robust. Further analysis would be needed to establish if the proposed clustering is strongly holding up statistically.

Given the identified group subdivision for episodes found, an interesting exploration is the possibility that there is any pattern related to the writer of the episode, implying that certain writers have a bias towards certain narrative segment structures. The visualization of the analysis in figure 24 suggest that the hypothesis is not supported, and is further explored by section **??**.

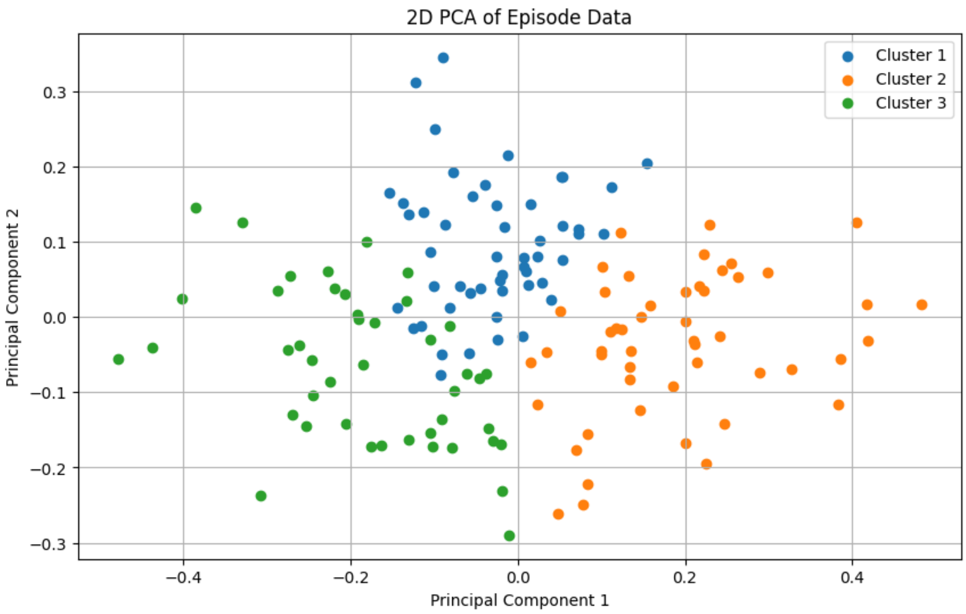

Fig. 23 Pca of the clustering in 2 dimensions

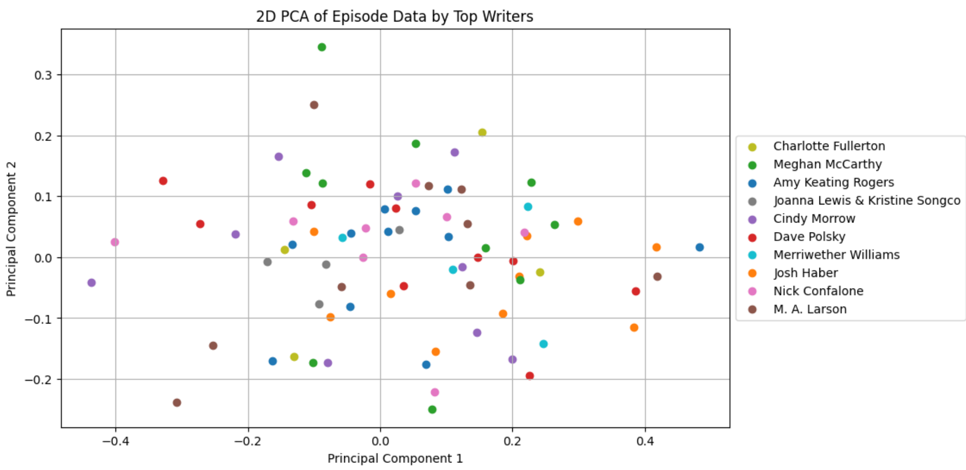

Fig. 24 Pca of the clustering in 2 dimensions, colors assigned to writers of the episodes

4.4 Incidence of writers over plot emotion distribution

Another quick exploration we made was to discover if certain writers were associated with certain emotions. Figure 25 suggests that the emotions in the episodes of all writers are similar, indicating a coherent style across the TV show, which is expected. No writer seems to predominantly be assigned to write in a certain emotional style over another.

Fig. 25 Breakdown of emotion distributions for characters during the episode "Yakity-Sax"