Text Based Emotion Detection

Shihab Hamati
Alfredo Funicello

26.01.2024

- **Objectives**
- Dataset
- Emotion Detection Models
  - Off-the-Shelf HuggingFace Transformer
  - Zero-Shot ChatGPT through Prompt Engineering
  - Task-Aware Fine-Tuned DistilBERT model
- Analyses
  - Overall Character Profile
  - Timeline across seasons and episodes
  - Writers' styles

Project Objectives

This project aims to:

- **A** explore modern **machine learning** tools in the context of text-based emotion detection
- **B** utilize **data science** analyses to understand better how characters exhibit their emotions in a children TV show.

- Objectives
- **Dataset**

Ⓐ Emotion Detection Models

  - Off-the-Shelf HuggingFace Transformer
  - Zero-Shot ChatGPT through Prompt Engineering
  - Task-Aware Fine-Tuned DistilBERT model

Ⓑ Analyses

  - Overall Character Profile
  - Timeline across seasons and episodes
  - Writers' styles

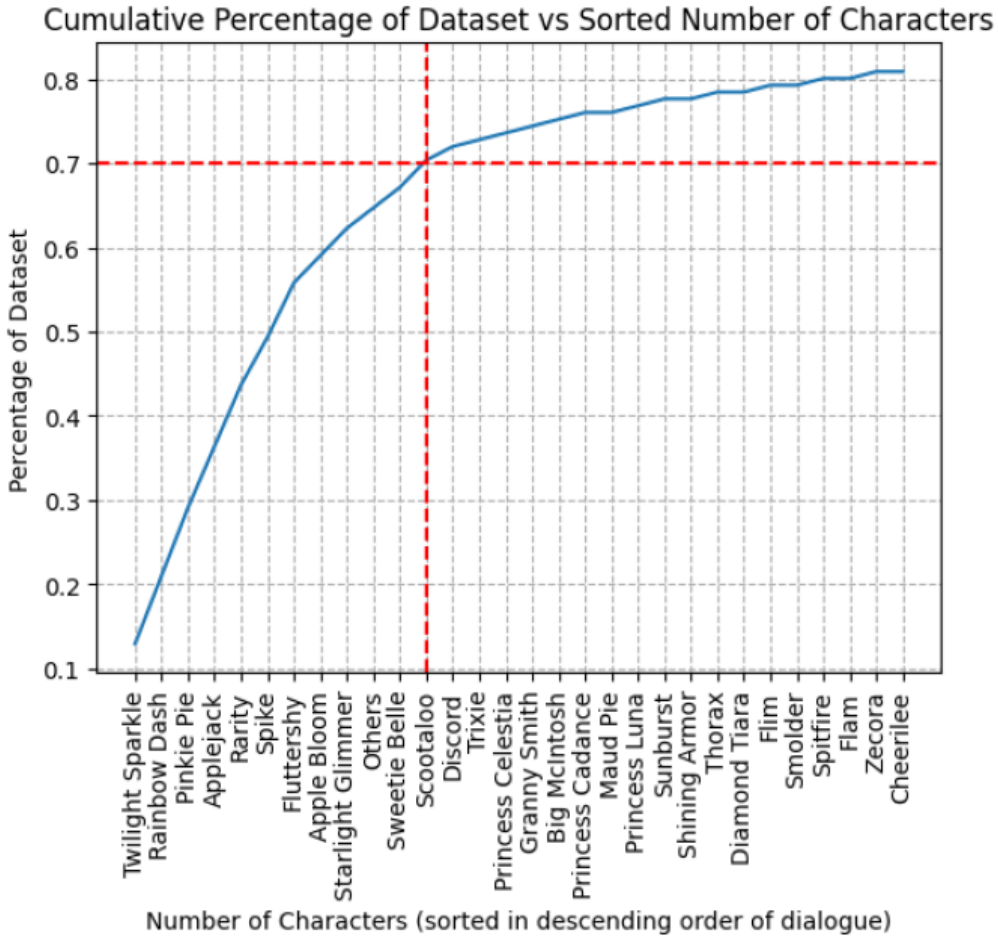
Dataset Description

- 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, with **36,859 rows of structured data**
- 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**

title	writer	pony	dialog
Friendship is Magic, part 1	Lauren Faust	Twilight Sparkle	and harmony has been maintained in equestria f...
Friendship is Magic, part 1	Lauren Faust	Twilight Sparkle	oh sorry girls i ve got a lot of studying to c...
Friendship is Magic, part 1	Lauren Faust	Twilight Sparkle	i know i ve heard of the elements of harmony

Dataset Preprocessing

- **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
- The graph of the cumulative count of datapoints vs characters was used to guide the data trimming decision; **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 25,419 dialogues



- Objectives
- Dataset

Ⓐ Emotion Detection Models

- Off-the-Shelf HuggingFace Transformer
- Zero-Shot ChatGPT through Prompt Engineering
- Task-Aware Fine-Tuned DistilBERT model

Ⓑ Analyses

- Overall Character Profile
- Timeline across seasons and episodes
- Writers' styles

Emotion Detection Models

3 different text-based emotion classification approaches were used :

❶ Off-the-Shelf Emotional Detection Transformer (**HuggingFace**)

❷ Zero-Shot LLM Emotion Labelling, through Prompt Engineering (**ChatGPT**)

❸ Task-aware Fine-Tuned Repurposed Pre-Trained LLM (**DistilBERT**, **ktrain**)

1 Off-the-Shelf Emotional Detection Transformer (from HuggingFace)

- 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**
- It uses **Ekman's 6 basic emotions** (anger, disgust, fear, joy, neutral, sadness, surprise) **plus neutral**
- **Pretrained on a balanced set of labelled datasets** containing around 20k data points and achieving a 66% accuracy
- It is **quite popular** with around 1.7 million downloads

1 Off-the-Shelf Emotional Detection Transformer (from HuggingFace)

- 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)

2 Zero-Shot LLM Emotion Labelling, through Prompt Engineering (ChatGPT)

- Advances in LLM-based chatbots allow for excellent **zero-shot or few-shot learning using generalized LLM models** for data augmentation
- 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 (1) a **system message used to deliver the task guidelines** and (2) a **human message used to deliver the dialog entry from the data**
- With an OpenAI **limit of 10k requests/day**, the labelling **spanned 3 days**

2 Zero-Shot LLM Emotion Labelling, through Prompt Engineering (ChatGPT)

- The **guideline message** used:

```
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.
```

- The **tiktoken** python library was employed to **estimate the cost** of OpenAI API usage for our chosen model, with a cost of $0.001/token

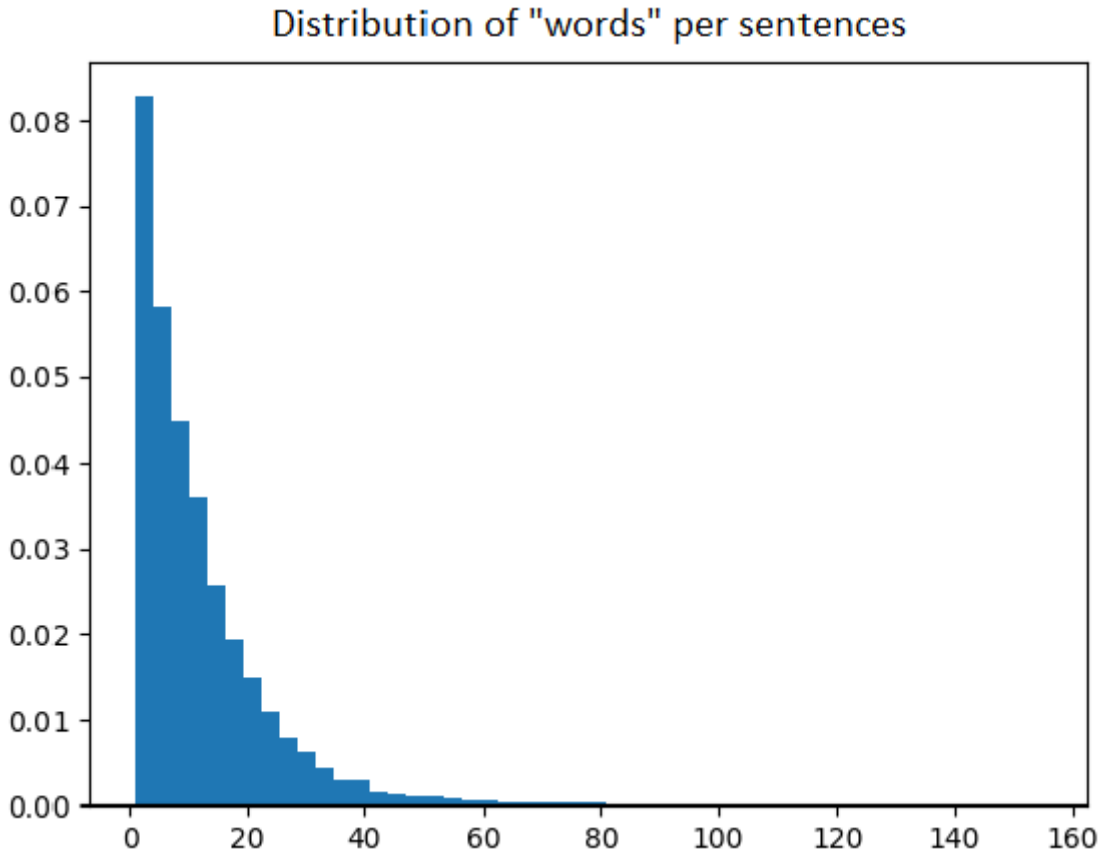
	Price	Percentage of total
Dialogue	$ 0.45	32.1 %
Guidelines	$ 0.94	67.9 %
Total	$ 1.39	

3 Task-Aware Fine-Tuned/Repurposed Pre-Trained DistilBERT model (ktrain)

- An alternative **more robust yet more involved** method is to fine tune a pretrained LLM neural net such as BERT or DistilBERT to our specific task
- The **ChatGPT obtained labels were used as a target column for a Text Classification task**; but any other labelled similar dataset could be used
- The **`ktrain` library 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

3 Task-Aware Fine-Tuned/Repurposed Pre-Trained DistilBERT model (ktrain)

- Also, a maximum length parameter of 100 was used, and this is reasonable since **almost the entire dataset falls into this constraint**

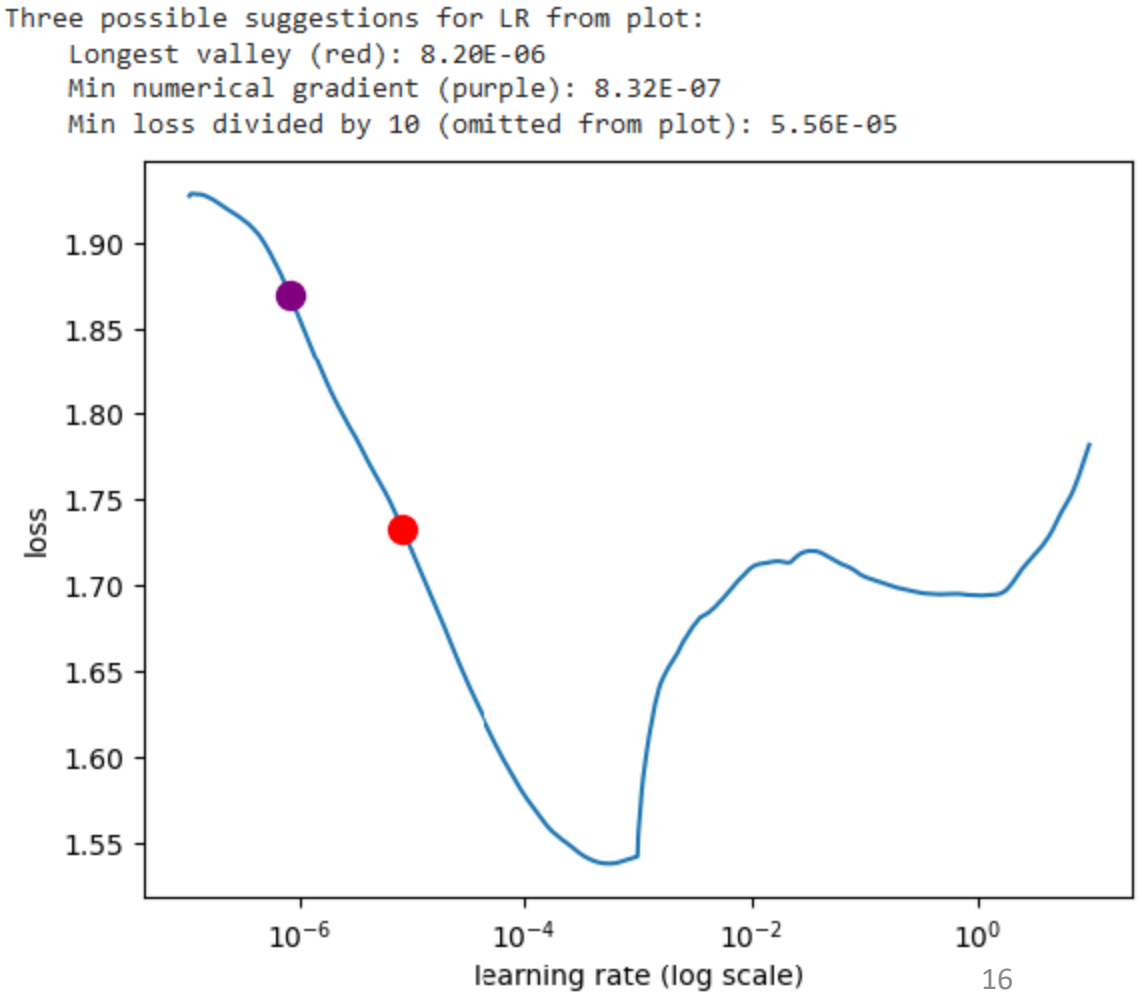


3 Task-Aware Fine-Tuned/Repurposed Pre-Trained DistilBERT model (ktrain)

- The first step was to **clean the data**, for example by omitting non-letter characters to 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%)

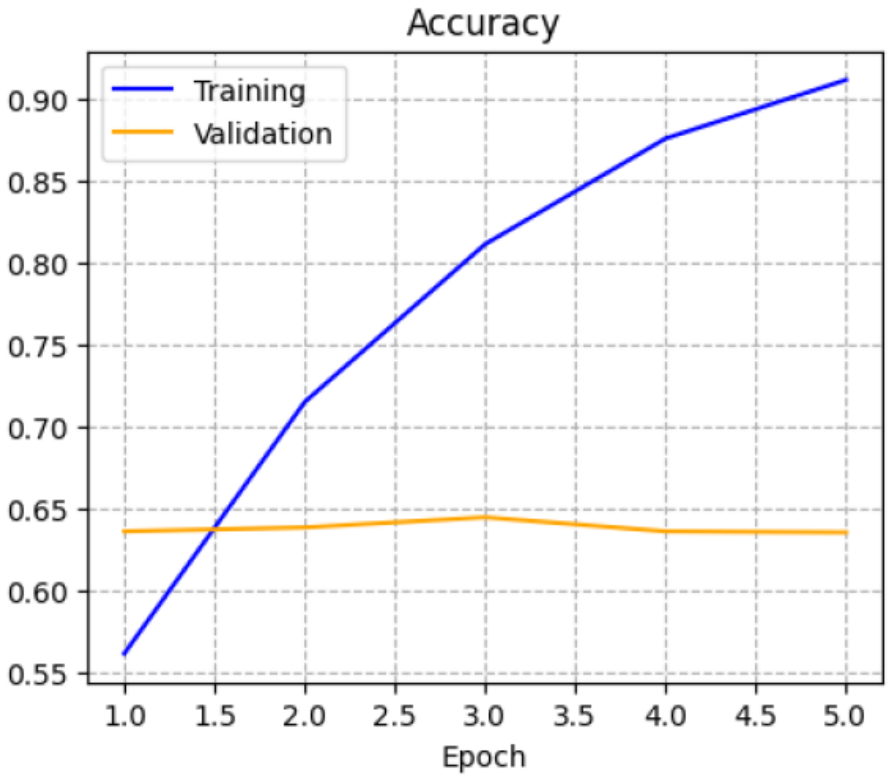
3 Task-Aware Fine-Tuned/Repurposed Pre-Trained DistilBERT model (ktrain)

- 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
- Optimal LR suggestions vary from 8.32e−7 to 5.56e−5. These have been **shown to provide ideal choices to pass to the fitting function** (used as an upper limit to the learning rate)



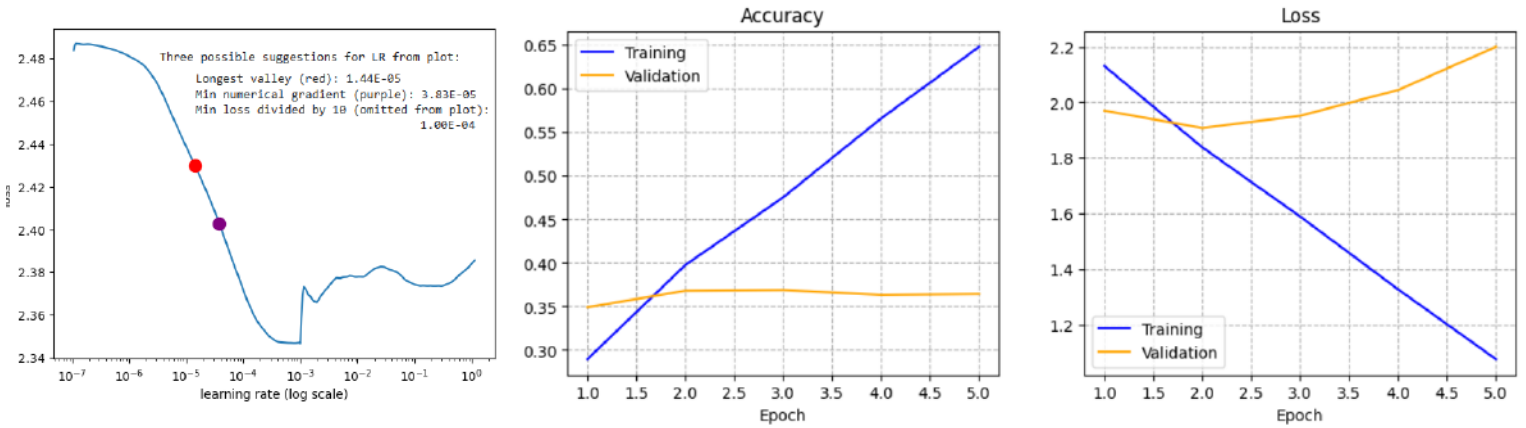
3 Task-Aware Fine-Tuned/Repurposed Pre-Trained DistilBERT model (ktrain)

- Since using a pretrained model **not a lot of time and resources needed to adjust to a new task**
- The training takes **approximately 4 mins/epoch**
- After training it for 5 epochs, **the model 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** (64%)
- In further steps, one might use a smaller model, increase the dataset size, or add counter measures **to reduce this over-fitting problem**



3 Task-Aware Fine-Tuned/Repurposed Pre-Trained DistilBERT model (ktrain)

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



- Objectives
- Dataset

Ⓐ Emotion Detection Models

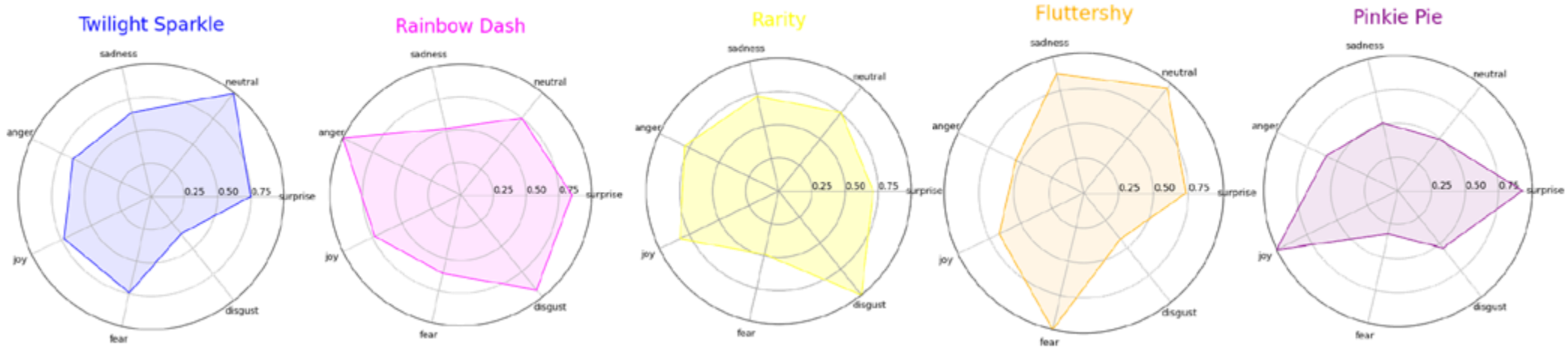
- Off-the-Shelf HuggingFace Transformer
- Zero-Shot ChatGPT through Prompt Engineering
- Task-Aware Fine-Tuned DistilBERT model

Ⓑ **Analyses**

- **Overall Character Profile**
- Timeline across seasons and episodes
- Writers’ styles

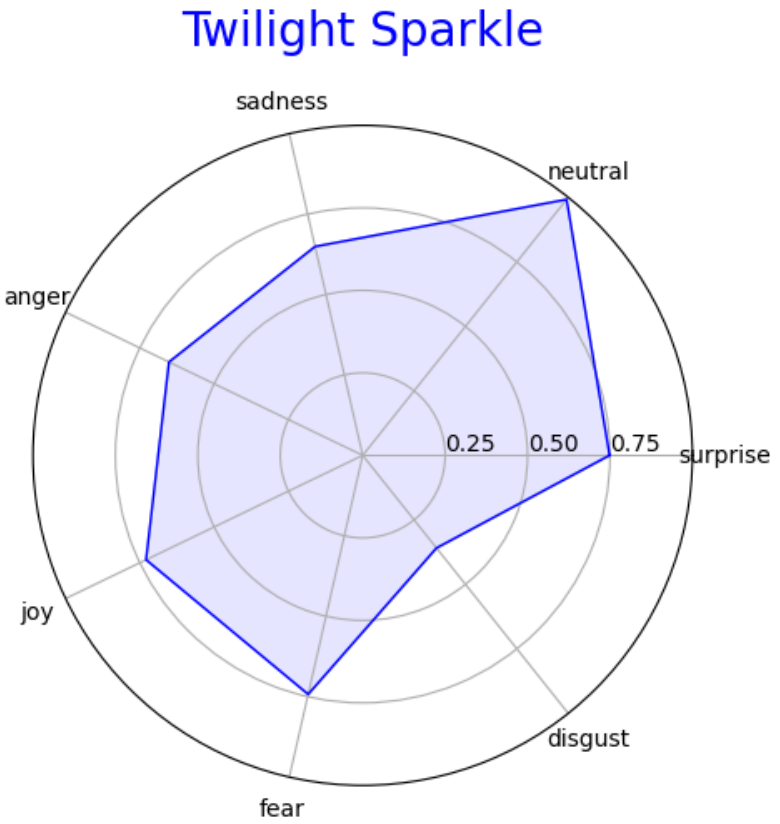
Overall Character Profile

- **Season 1** is where characters get introduced and are most well characterized to get viewers accustomed to the different personalities
- The distribution of emotions for each of them has been calculated and then **normalized on a 0-1 scale** using the max values over all characters
- How well this methodology describes the character will be explored by **comparing the radar charts to the Wikia pages** of each of the characters



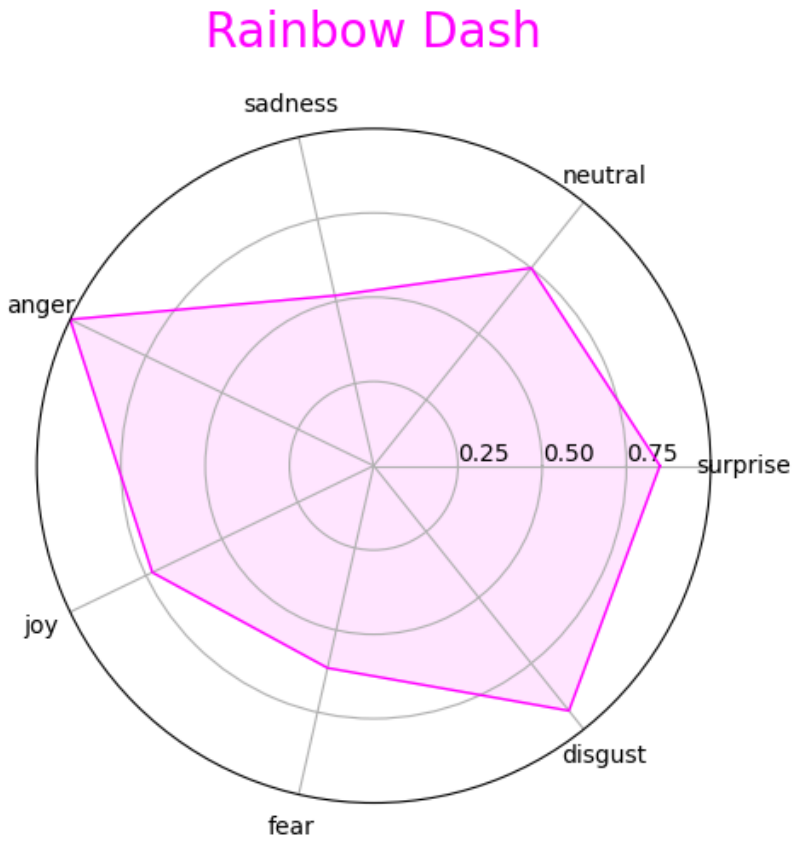
Character Profile: Twilight Sparkle

- 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**
- **Wikia:** *She tries to be rational in unfamiliar situations, Twilight tends to be skeptical of unproven claims; however, Twilight can lose her cool under stress*



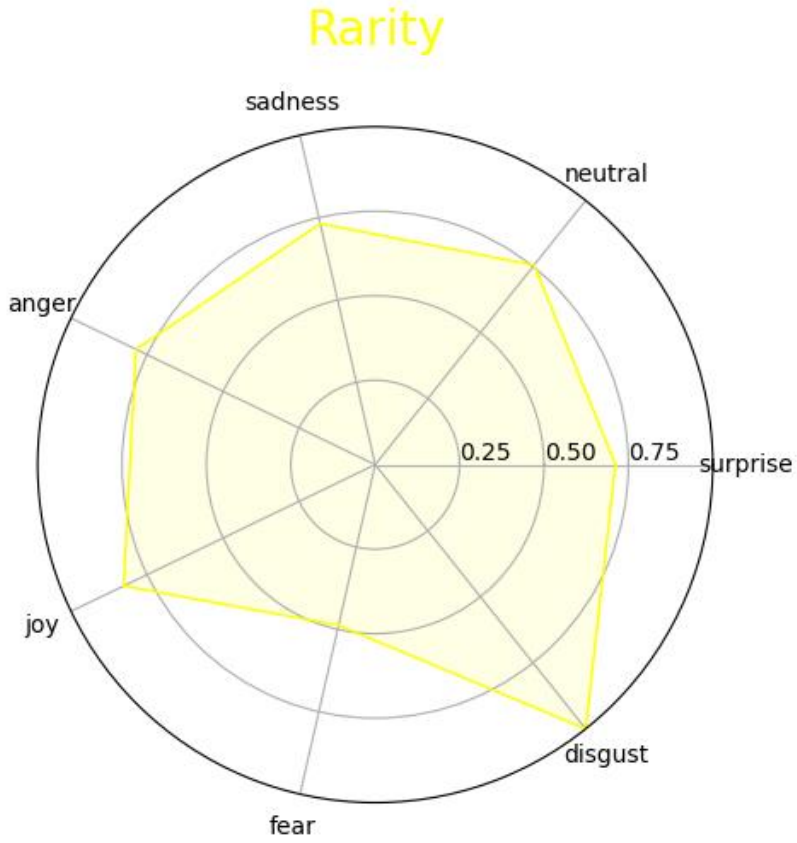
Character Profile: Rainbow Dash

- 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**



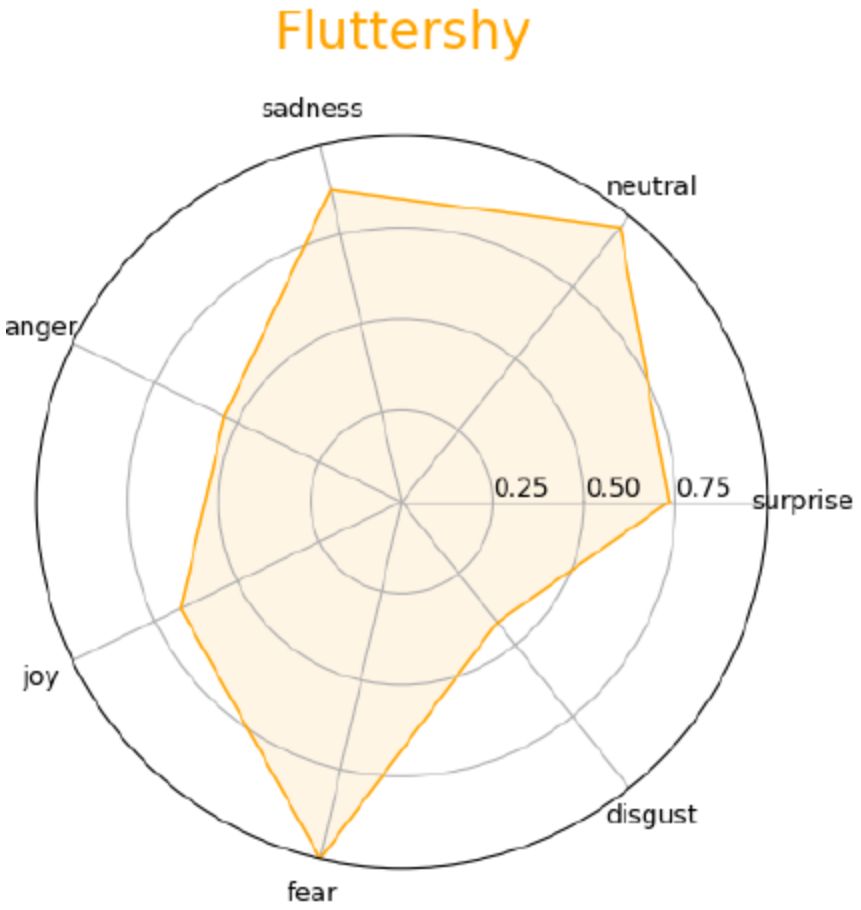
Character Profile: Rarity

- **Wikia:** *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*
- 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**



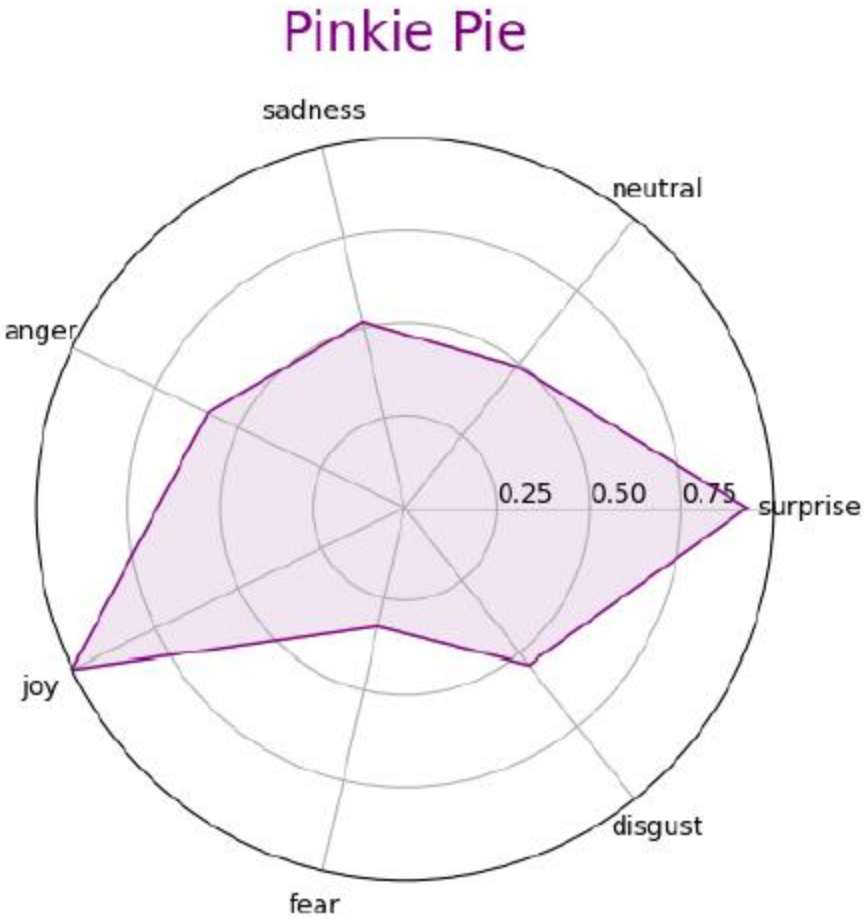
Character Profile: Fluttershy

- **Wikia:** *When Fluttershy is first introduced in the series, she barely manages to tell Twilight Sparkle her own name on account of her timidity*
- In her emotions distribution, she is the **most fearful and is often sad** which aligns with her Wikia



Character Profile: Pinkie Pie

- **Wikia:** *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 as she is the character **expressing the most joy and displays a high amount of surprised dialogue**



- Objectives
- Dataset

Ⓐ Emotion Detection Models

- Off-the-Shelf HuggingFace Transformer
- Zero-Shot ChatGPT through Prompt Engineering
- Task-Aware Fine-Tuned DistilBERT model

Ⓑ Analyses

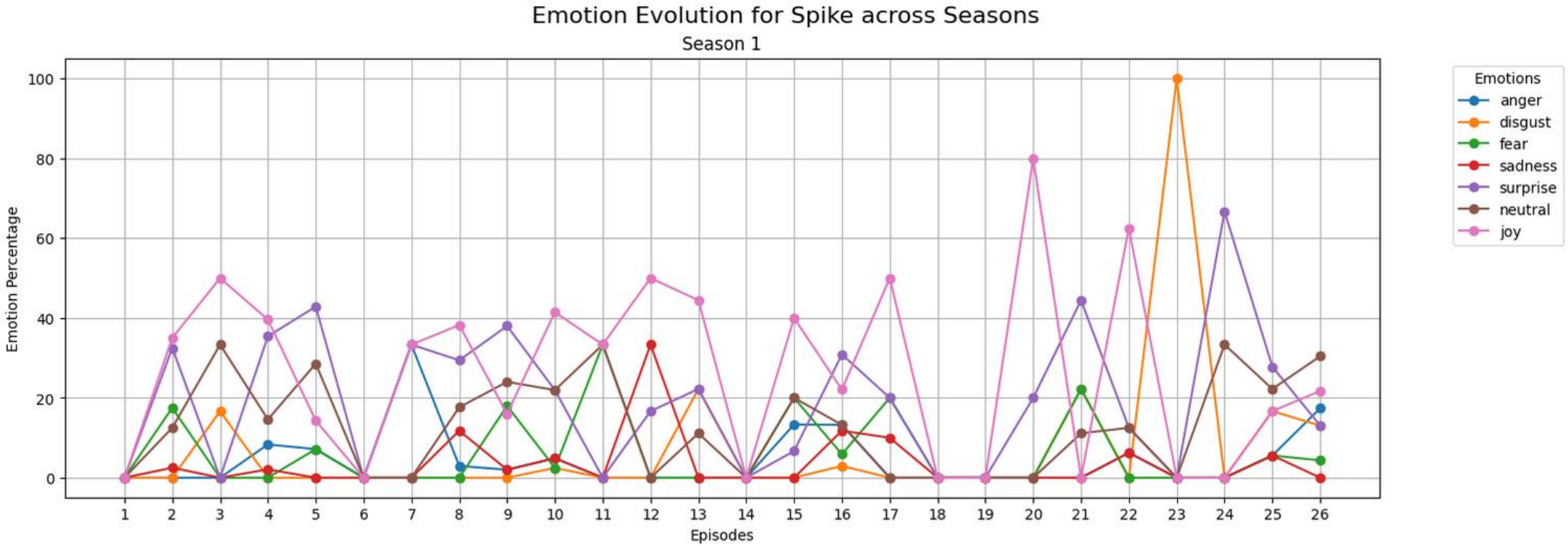
- Overall Character Profile
- **Timeline across seasons and episodes**
- Writers' styles

Emotional timeline across seasons and episodes

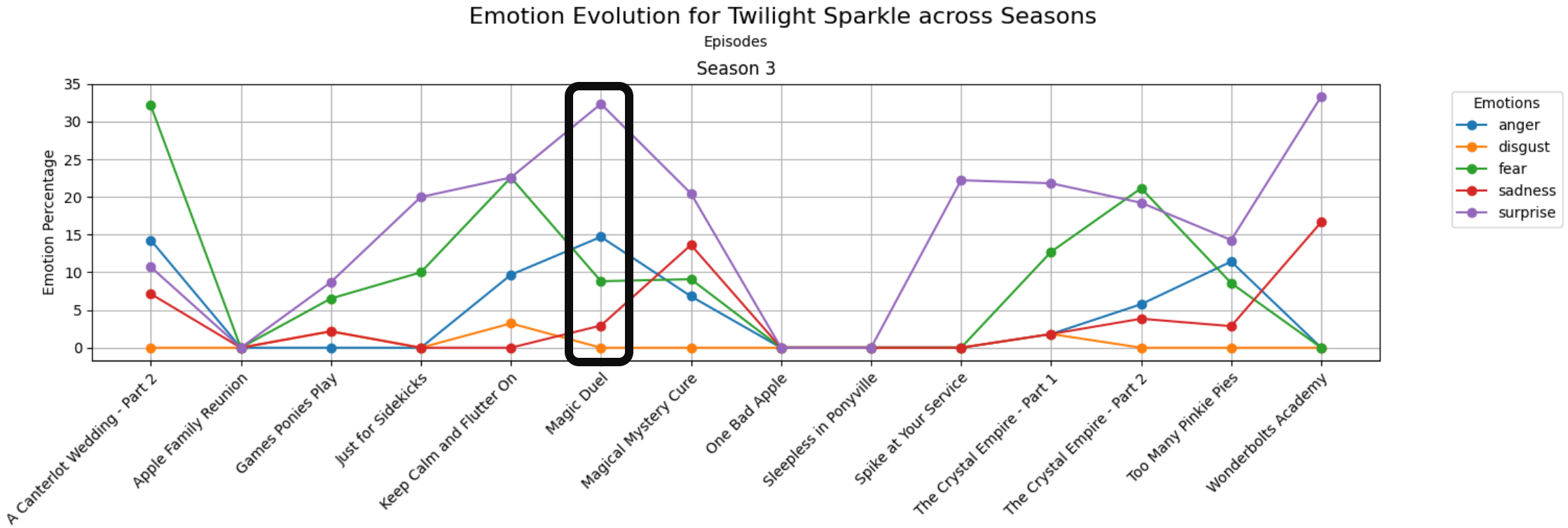
```
{
  {
   'episode_name1':
    {
      'character_name1':
       {
         'anger': 2.5,
         'disgust': 0.0,
         'fear': 5.0,
         'joy': 40.0,
         'neutral': 30.0,
         'sadness': 15.0,
         'surprise': 7.5
       },
     ...
    },
    'season': 1
  }
  ...
}
```

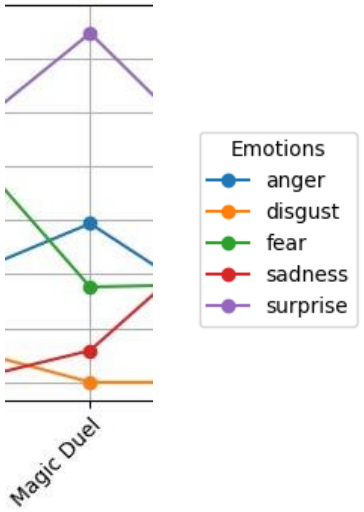
- Constructed a **data structure storing the percentage of emotional labels** found in each episode's dialogue for each character
- Analytical approach to **quantify and track the emotional distribution** throughout the episodes and seasons.
- Interest in **Macro-level emotional patterns over seasons** and **Micro-level patterns over episodes**

- **No relevant Macro-level pattern** of emotional evolution when between seasons
- Might indicate that **episodes are unrelated** and a season doesn't generally follow any predefined plot scheme.
- **Predominance of Joy** in the characters' utterances aligns well with the **show's audience being kids** and the central theme of friendship

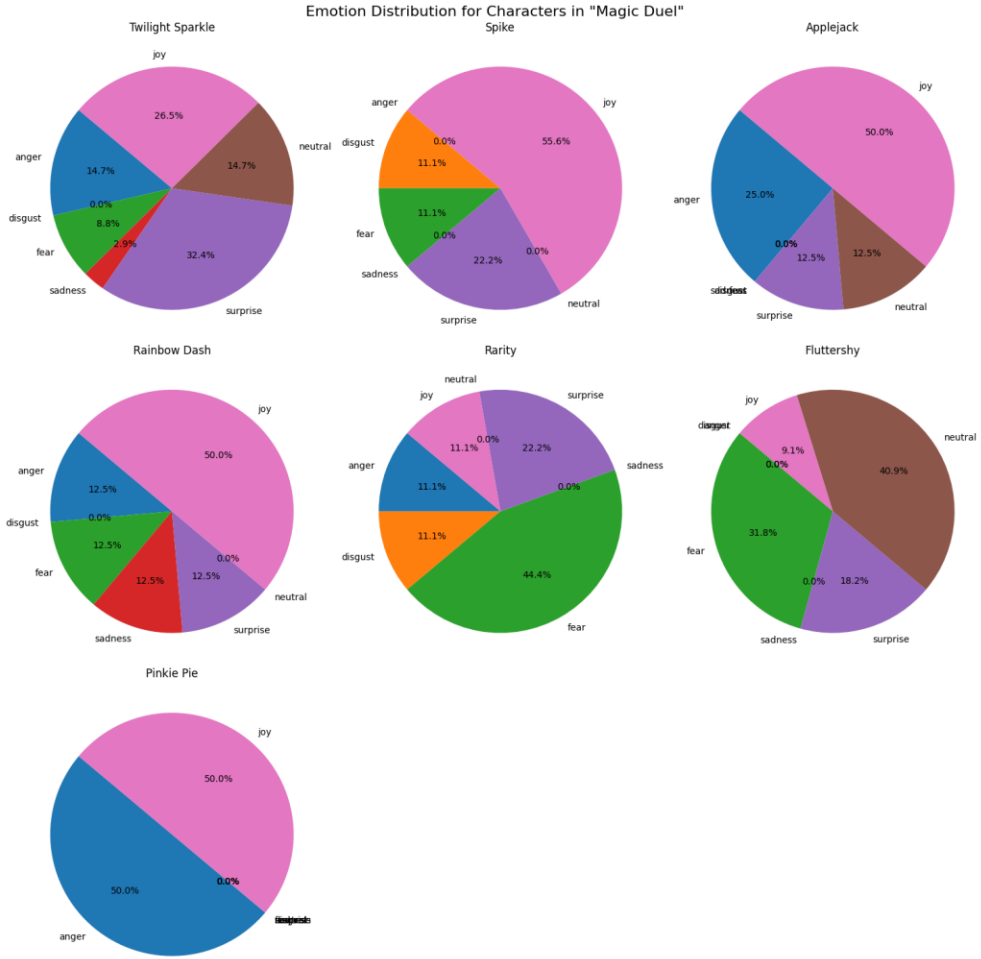


- **<u>Micro-level analysis</u>** (excluding joy, neutral) highlights episodes featuring heightened emotional experiences for the characters
- In episode '*Magic duel*' big peak of **surprise** suggests **astonishment** or **amazement**, accompanied by peaks of **fear**, **anger** and **sadness** suggest **shock** and **dismay**; negative setting of the episode

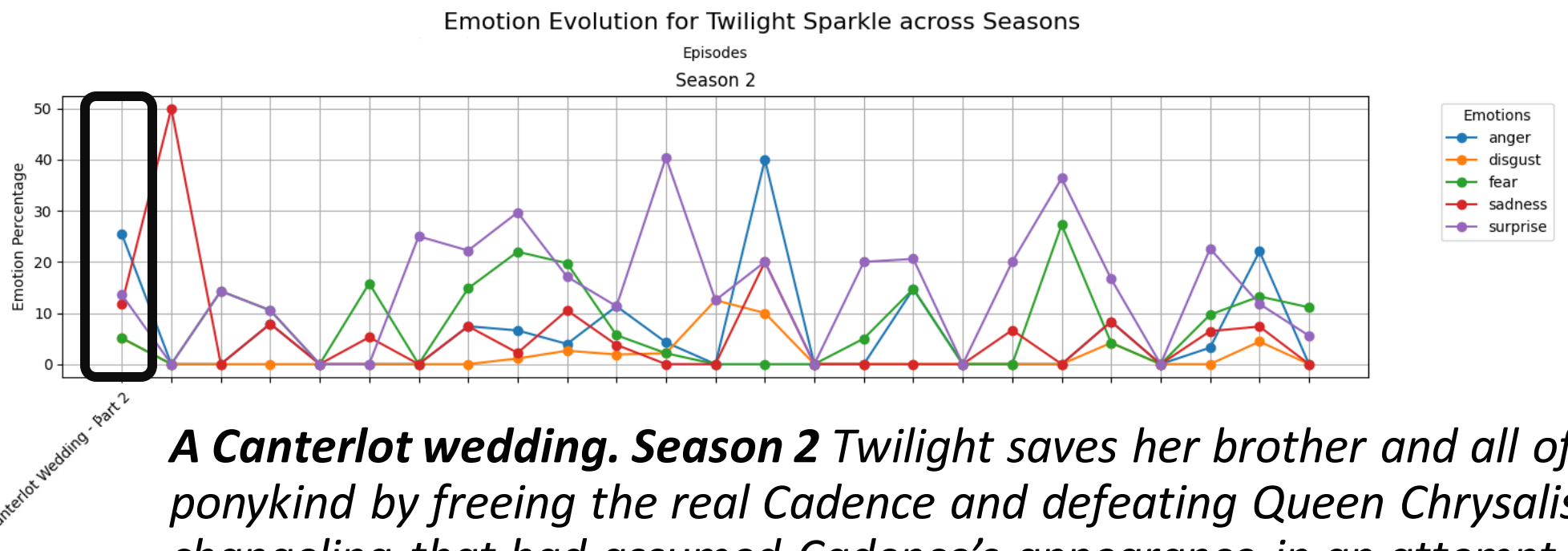




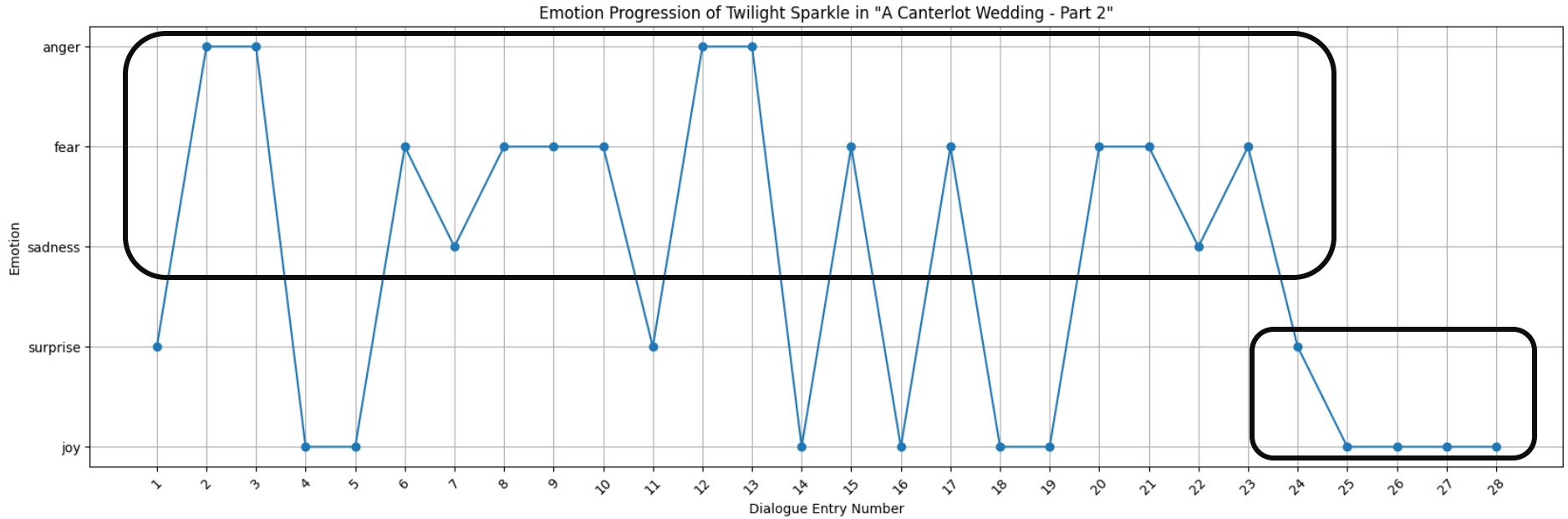
Magic Duel. Season 3 *Seeking revenge the powerful Trixie returns to Ponyville, defeating Twilight in a magic duel. She exiles Twilight from Ponyville, who must figure out a way to best Trixie if she is to return home.*



- Episode hosts an **antagonist**.
- Losing against Trixie and being exiled could have been the significant trigger event for Twilight's shock
- Most of the cast expresses out of the average levels of **fear** and **anger**
- Even in negative episodes the language tends to be **joyful**

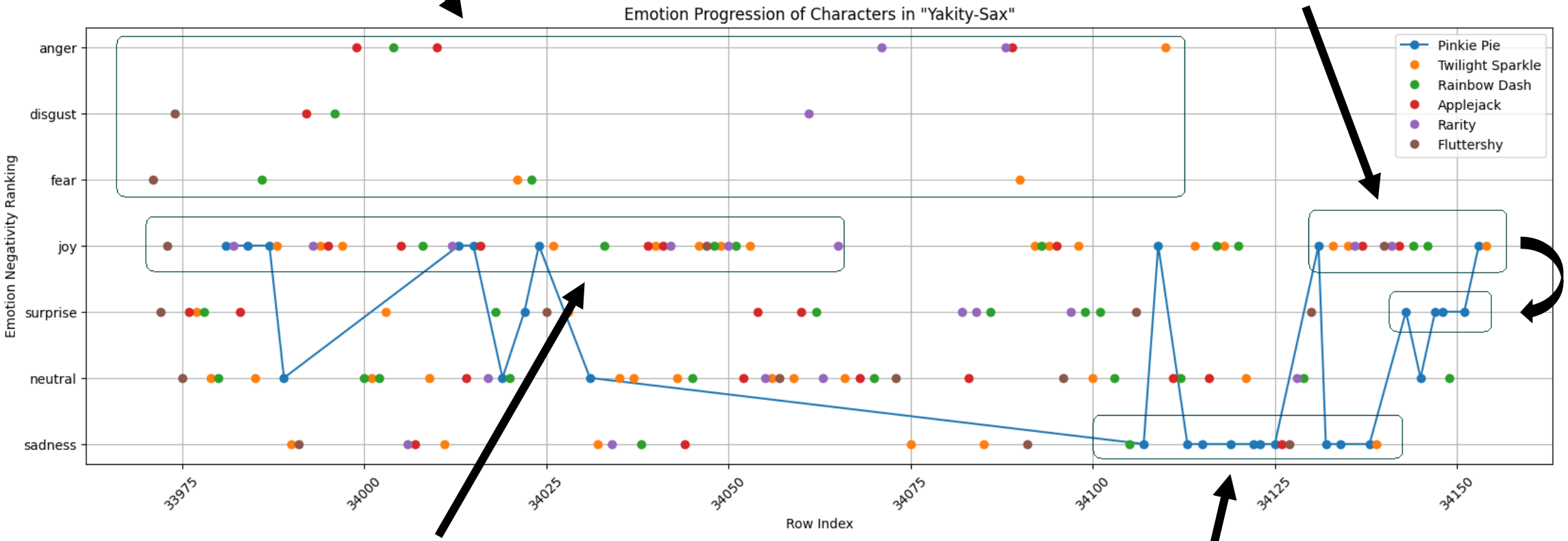


A Canterlot wedding. Season 2 *Twilight saves her brother and all of the ponykind by freeing the real Cadence and defeating Queen Chrysalis, a changeling that had assumed Cadence's appearance in an attempt to take over Equestria.*



1. *{...} Unfortunately, Pinkie's playing causes constant disruption for her friends' daily activities. {...}*

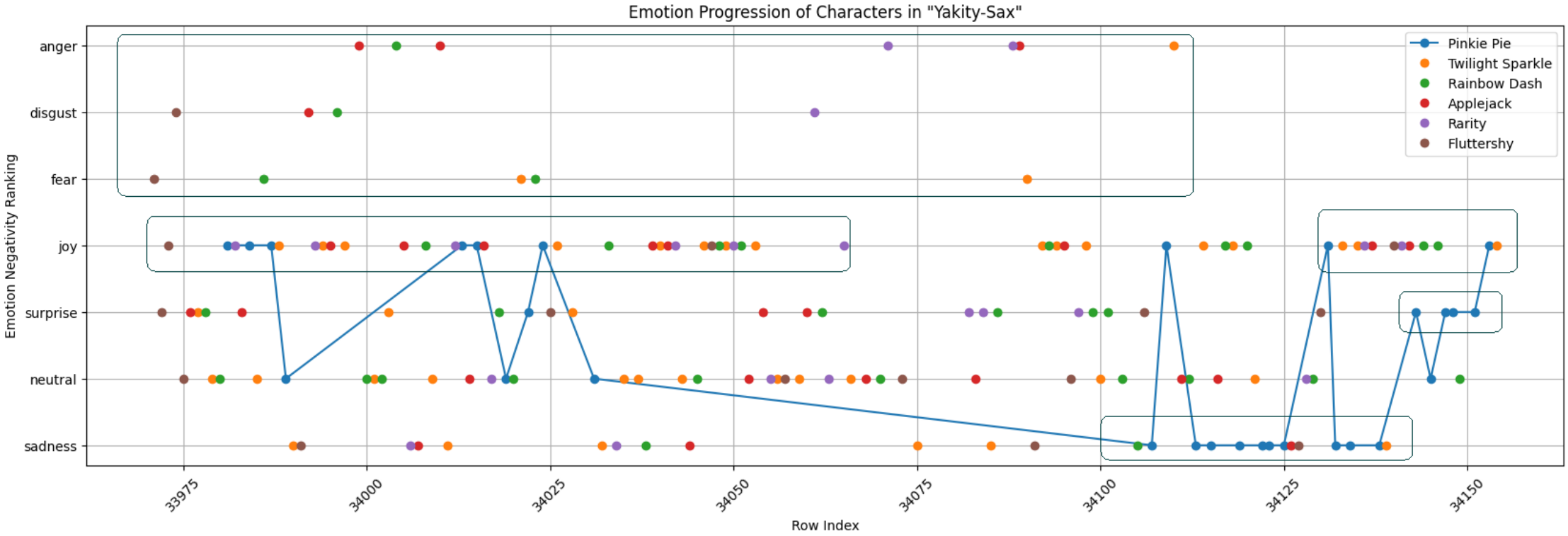
4. *{...} Twilight and the others realize they were wrong to make Pinkie stop doing something she enjoys, they encourage her to continue playing as long as it makes her happy {...}*



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

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

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*



Reagan et al. *EPJ Data Science* (2016) 5:31
DOI 10.1140/epjds/s13688-016-0093-1

EPJ.org

EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE Open Access

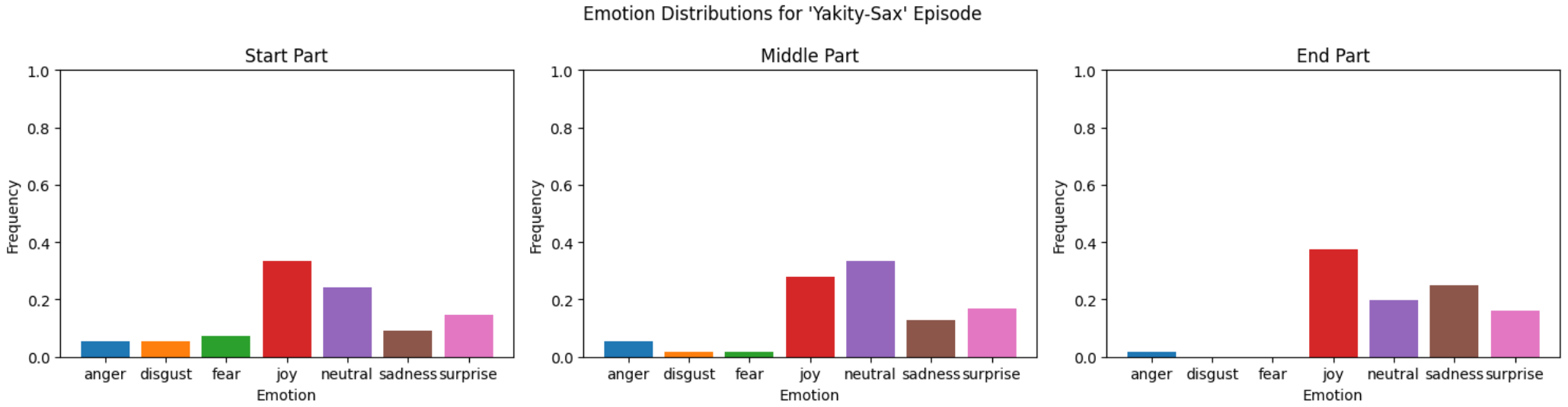
The emotional arcs of stories are dominated by six basic shapes

Andrew J Reagan[1*], Lewis Mitchell[2], Dilan Kiley[1], Christopher M Danforth[1] and Peter Sheridan Dodds[1]

Researchers took the emotional arcs of 1300+ novels from Project Gutenberg, used modern tech to **analyze the emotional arcs**, and then **identified 6 patterns seen over and over again** in western storytelling.

1. Rags to Riches (**rise**)
2. Riches to Rags (**fall**)
3. Man in a Hole (**fall** then **rise**)
4. Icarus (**rise** then **fall**)
5. Cinderella (**rise** then **fall** then **rise**)
6. Oedipus (**fall** then **rise** then **fall**)

- The previous analysis suggests that some episodes are structured with a **clear sequence of events** that shape the story's progression
- *Hypotesis*: there might be a **pattern in how emotions are distributed** across the timeline of the episode
- *Methodology*: **subdividing episodes in 3 segments** (start, middle and end) of equal size, finding emotion distributions

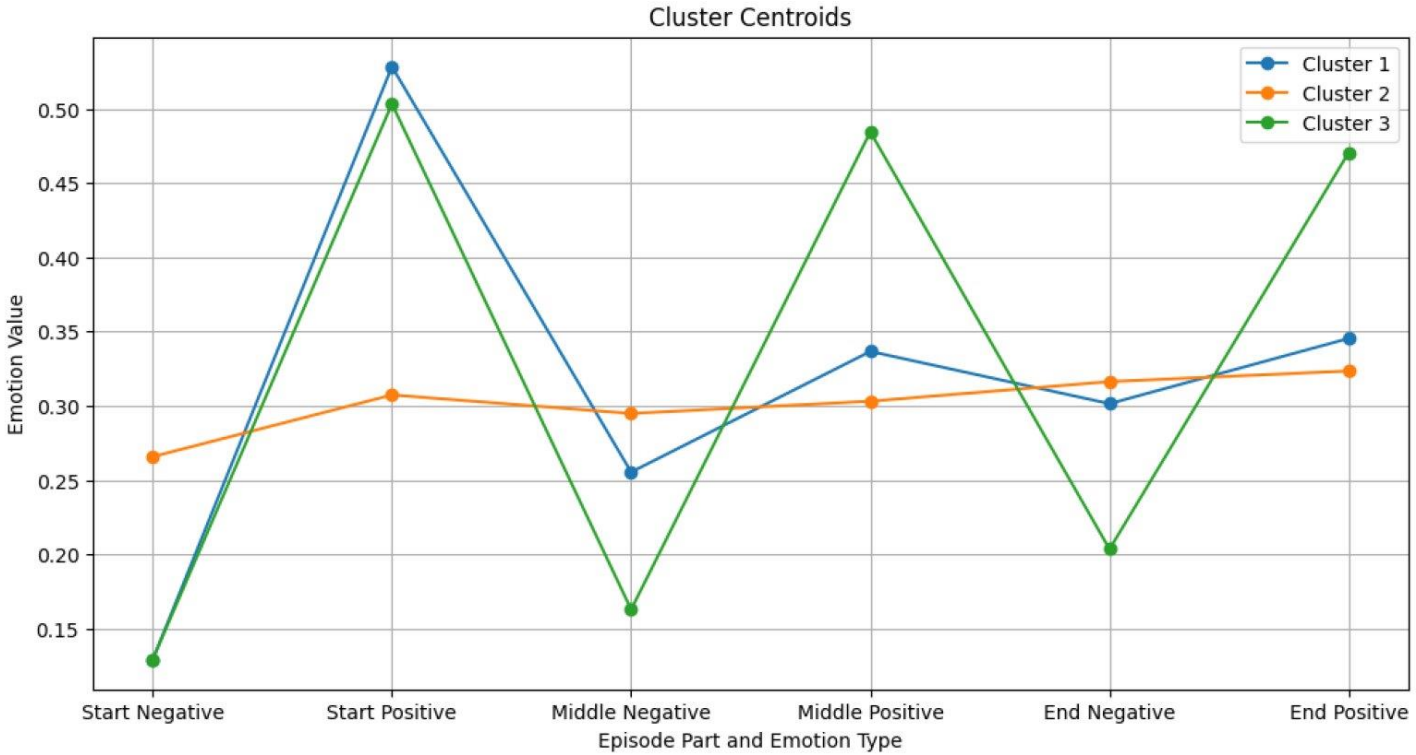


- **High variance** of distributions across the segments found
- **Binning** the emotions in **negatives** (anger, disgust, fear) and **positives** (joy)
- *Assumption*: plot segments can be generally correctly simplified to negative atmosphere segments and positive atmosphere segments

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

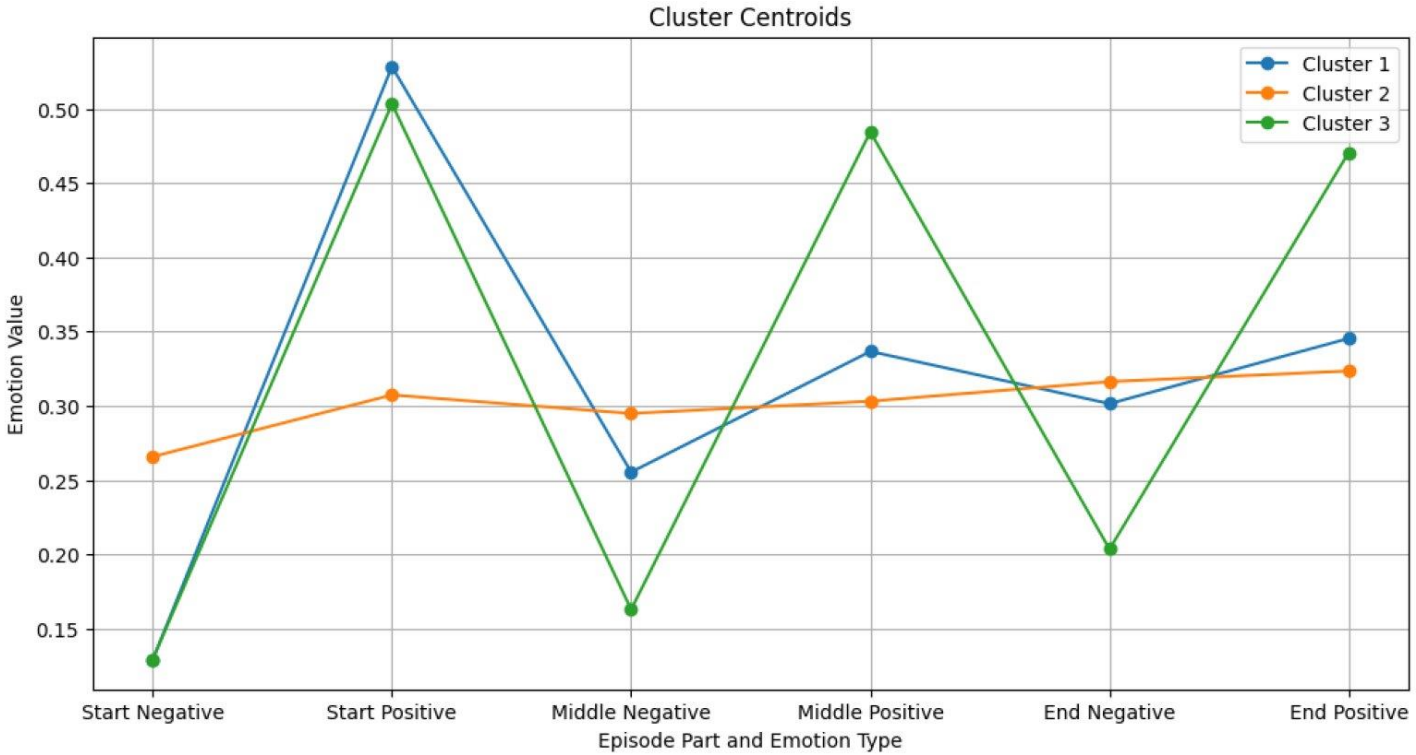
Segment	Negative %	Positive %
Start	0.176 ± 0.006	0.444 ± 0.0178
Middle	0.240 ± 0.009	0.369 ± 0.013
End	0.276 ± 0.010	0.375 ± 0.011

- **Start**: episodes on an overwhelmingly positive note.
- **Middle**: balanced
- **End**: part most emotionally charged



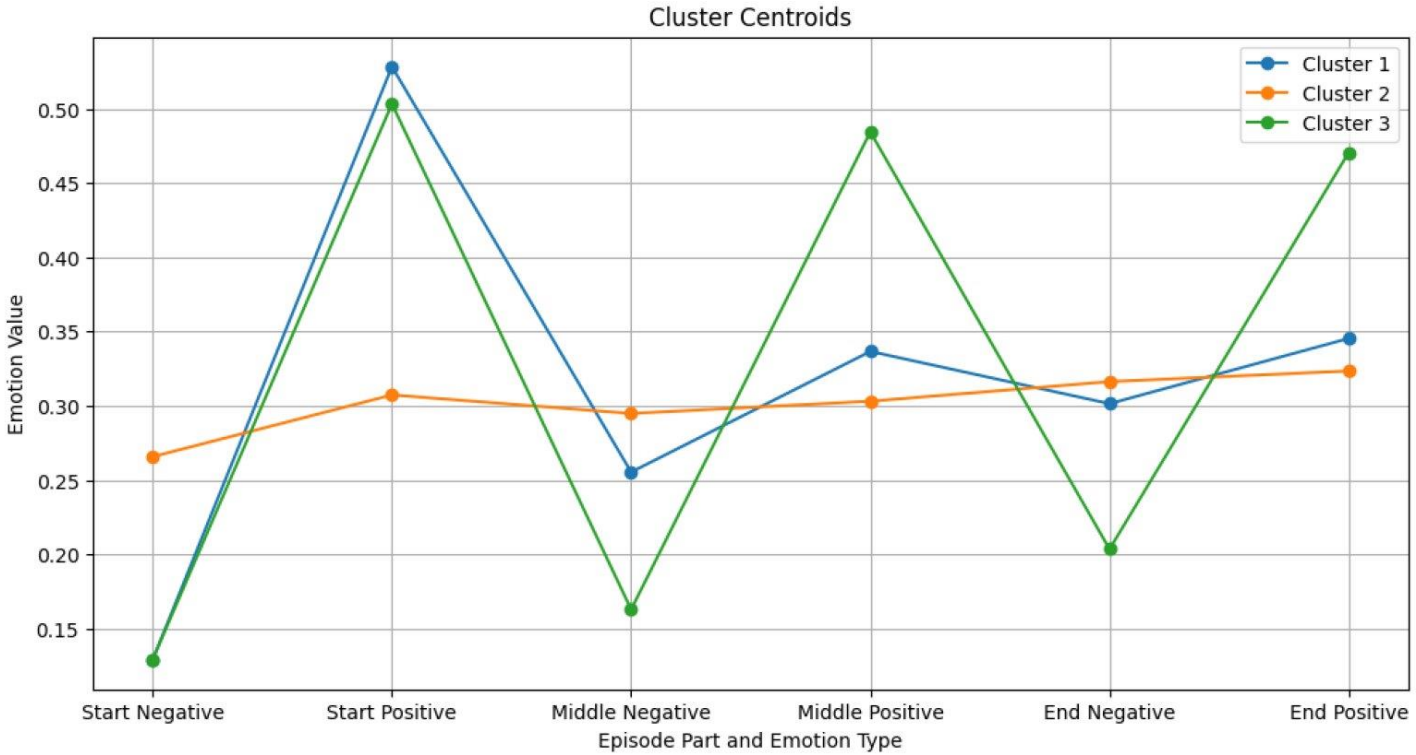
Cluster 3 is overwhelmingly positive episodes

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.*



Cluster 2 is episodes where there's an equal balance between positive and negative from the start

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.*



Cluster 1 is episodes that start positively, then something happens that introduces some negativity (everything was very fine then came the bad guy trope)

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.*

- Objectives
- Dataset

Ⓐ Emotion Detection Models

- Off-the-Shelf HuggingFace Transformer
- Zero-Shot ChatGPT through Prompt Engineering
- Task-Aware Fine-Tuned DistilBERT model

Ⓑ **Analyses**

- Overall Character Profile
- Timeline across seasons and episodes
- **Writers' styles**

Writers' Effect on Emotions in Episodes

- Given the identified group subdivision for episodes found, **an interesting exploration is any pattern related to writers**, implying that certain writers have a bias towards certain narrative segment structures.
- The the previous analysis suggest that **this hypothesis is not supported**

Writers' Effect on Emotions in Episodes

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