Storify: Music Suggestion Model for Social Media Stories

Sanjana IIITD, Hardi Parikh, Vishnu Mothukuri, Shlok Mehroliya, Om Mehroliya, Rahul Ajith

INTRODUCTION

In the digital storytelling realm, integrating music into social media stories significantly boosts the emotional and aesthetic appeal of user-generated content across platforms such as Instagram, Facebook, and WhatsApp. Despite its crucial role in enhancing narratives, users often encounter difficulties in selecting music that harmoniously complements their visual stories, directly impacting viewer engagement and satisfaction.

Addressing this, the Storify model emerges as a solution aimed at automating the selection of background music by meticulously analyzing the visual content. The challenge primarily lies in the users' struggle to find music that aligns with the mood and theme of their images, stemming from the intricate process of interpreting visual themes and matching them with appropriate musical genres.

To address this challenge, our Storify model adopts a sophisticated approach that hinges on the synergy between advanced emotion detection and music recommendation technologies. Utilizing a state-of-the-art emotion detection model, Storify accurately categorizes user-generated content into one of seven distinct emotional tags. This classification is pivotal in understanding the underlying mood conveyed by the visual stories. Building on this emotional foundation, Storify then leverages the Spotify API to curate a personalized playlist, meticulously selected to resonate with the detected emotion.

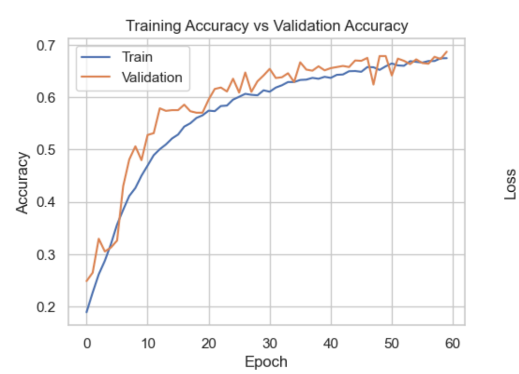
UPDATED BASELINE RESULTS

System prototype overview:

The original version of the Storify model was designed as a pioneering tool to assist users in selecting background music that is in harmony with the visual and emotional content of their social media stories. It harnessed the capabilities of deep learning to interpret the visual elements and natural language processing (NLP) to parse the accompanying captions. The goal was to seamlessly match these narratives with the appropriate musical tracks, thereby elevating the storytelling experience on social media platforms like Instagram, Facebook, and WhatsApp.

In our quest for refinement, we have implemented substantial updates to the Storify model, which have resulted in marked improvements in both accuracy and performance. The graphs shown illustrate the trajectory of our system's enhanced learning capability over numerous training epochs.

- **Training Accuracy vs. Validation Accuracy:** Our updated model demonstrates a convergence of training and validation accuracy, highlighting the effectiveness of our enhancements. The accuracy shows a significant and consistent upward trend, peaking at around 67%, which indicates a robust learning from the training data without substantial overfitting.
- **Training Loss vs. Validation Loss:** Correspondingly, the training and validation loss metrics depict a steep decline in the initial epochs, leveling off as the model approaches an optimal state. This indicates an improved generalization in recognizing and interpreting the emotional context of visual stories.



Performance Metrics:

Through rigorous testing and refinement, we have achieved substantial enhancements in the system's accuracy and efficiency.Comparative Analysis of Accuracy or Other Relevant Metrics Before and After Updates :

- The initial prototype of the Storify model provided a solid foundation with a baseline accuracy that demonstrated the feasibility of our approach. However, it became evident that there was significant room for improvement, particularly in terms of the system's ability to generalize across diverse data inputs.
- **Accuracy:** Initially, our model's accuracy hovered around the 40% mark, a modest start for complex emotional recognition tasks. Post-updates, we've observed a remarkable jump in accuracy to approximately 67%, indicating a much more adept system at correctly identifying and tagging emotions based on visual and textual inputs.
- **Loss Metrics:** The initial average loss for our model was quite high, suggesting a need for refinement in the model's ability to minimize the error rate during training. Through optimization techniques and enhanced training protocols, we have successfully brought down the loss metrics significantly. The validation loss now stabilizes at a much lower level, indicating a better fit of the model to the data.

Dataset used for testing:

The Storify model's performance is evaluated using a comprehensive dataset well-suited for our emotion detection and music recommendation tasks. We have utilized the widely recognized **FER-2013 dataset**, which is specifically designed for facial expression recognition tasks.

- **FER-2013 Dataset:** The Facial Expression Recognition 2013 (FER-2013) dataset is composed of 35,887 grayscale.The dataset is categorized into seven facial expressions that correspond to emotions: anger, disgust, fear, happiness, neutral, sadness, and surprise. This diversity in emotional expressions allows our model to accurately discern and tag the emotional context of social media content, forming the basis for our music recommendation system.
- **Preprocessing:**

Grayscale Normalization: Given that the images are already in grayscale, we normalize the pixel values to the range [0,1], improving the numerical stability of the neural network.

Image Augmentation: To increase the diversity of the dataset and prevent overfitting, we introduce image augmentation techniques such as rotation, width and height shifts, zooming, and horizontal flipping.

Resizing Images: Although the images are provided at a standard size, we resize them to fit the input requirements of our neural network architecture without compromising the integrity of the visual information.

Facial Detection and Alignment: We apply facial detection algorithms to ensure that faces in the images are centered and aligned, which is particularly important for accurate emotion detection.

PROPOSED METHODS FOR PROBLEM SOLVING

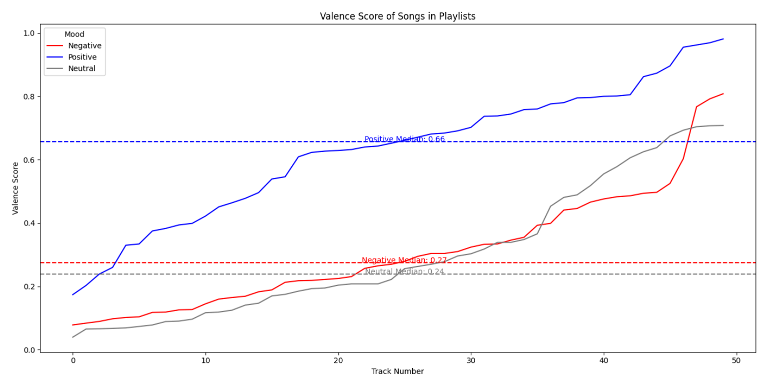
Implementations

- **Advanced Emotion Recognition:** Utilizing the latest developments in convolutional neural networks (CNNs) and recurrent neural networks (RNNs), our method will introduce refined layers and weight optimization techniques to better interpret the subtle nuances of human expressions.
- **NLP Augmentation:** We plan to incorporate state-of-the-art natural language processing advancements to enhance the interpretation of captions, focusing on semantic analysis to capture the mood and thematic elements more accurately.
- **Valence-Aware Music Matching:** By introducing valence as a parameter in our music selection algorithm, we ensure that the dynamic range of emotions is matched with music that has a corresponding energy level, whether it's uplifting, somber, or anywhere in between.
- **Hybrid Recommendation System:** We propose a hybrid system that combines content-based and collaborative filtering mechanisms from the Spotify API, allowing for more personalized and contextually relevant music recommendations.

Outcomes:

- **Enhanced Accuracy in Emotion Tagging:** With the improved emotion recognition features, we expect to see a rise in the accuracy of emotional tagging, thus ensuring that the selected music is a true reflection of the visual narrative's intended mood.
- **Greater Diversity in Music Selection:** Our advanced data analysis techniques will enable us to tap into a broader spectrum of musical genres and styles, catering to a wider array of user preferences and enhancing the personalization of the storytelling experience.
- **Improved User Engagement:** By addressing the limitations in the baseline results, such as the occasional mismatch in music and content mood, we anticipate a marked increase in user engagement and satisfaction.

Valence relation with songs selection:



Negative emotions

Positive emotions

Neutral emotions

The graph presents the valence scores of songs in playlists, sorted by mood: negative, positive, and neutral. It indicates that tracks in the positive playlist have a higher median valence score of 0.66, showing a trend towards more upbeat and cheerful music. Conversely, the negative mood playlist has a lower median valence score of 0.27, aligning with more somber or subdued tracks. The neutral playlist sits at the median valence score of 0.24, suggesting a balanced, neither particularly upbeat nor downcast, musical selection.

RESULT

Image Input :

```
1/1 [==============================] - 0s 62ms/step
Predicted Emotion: Happy
```

Final Output :

```
Songs for Happy Emotion Tag:
           artist                                  track_name  valence
      Chuck Berry                      No Particular Place To Go    0.980
         Four Tops   I Can't Help Myself (Sugar Pie, Honey Bunch)    0.971
      Jean Knight                                   Mr. Big Stuff    0.971
 Sharon Jones & The Dap-Kings           How Long Do I Have to Wait for You?    0.969
      Ann Peebles                        I Can't Stand the Rain    0.966
    Arthur Conley                              Sweet Soul Music    0.965
      Eddie Floyd                                 Knock on Wood    0.964
    Fontella Bass                                     Rescue Me    0.963
  Various Artists                      Respect - 2003 Remaster    0.961
   The Isley Brothers      This Old Heart Of Mine (Is Weak For You)    0.953
```

Top K songs are retrieved from spotify API based on the valence scores