Lionel Quintanilla, Andrew De La Rosa, Pouya Tavakoli, Brian Tran, Matthew Zaldana

Professor Maryam Qudrat

ENGR 361

30 November 2021

A Machine Learning Approach to Closed Captioning

For our research hypothesis, we want to create an application that uses machine learning techniques to produce closed captions on any video player with accuracy that exceeds current methods. We plan to achieve this by exploring two specific aims: first, by devising a machine learning technique that can create captions with a reasonable degree of accuracy, and secondly by creating a user-facing application that can take in video inputs and display captions by overlaying of these videos.

The issue of improving the accuracy of closed captioning is imperative because it intersects both legal and moral considerations. The main necessity for accurate captions comes from the fact that closed captions are legally required in the United States and other countries. However, deaf and hard of hearing people who rely on captions to understand video media are being underserved because current methods of captioning struggle with accuracy and readability. While there is research being done to improve the current methods, we want to take a novel new approach and use machine learning to bypass the issues with the current methods.

To research our hypothesis, we will conduct two separate experiments based on our specific aims. For the first experiment, we will create a list of three to six machine learning algorithms for creating closed captions. We will then set up each algorithm and on a dedicated computer and have it caption a series of datasets containing images and videos. We will repeat this process for every algorithm so that their results can be compared. To analyze the results, take the captions generated by the algorithms and run them through several tests for natural language to determine the accuracy of the captions. The accuracy scores will then be averaged for each technique to compare them. For the second experiment, we will gather a focus group of deaf/hard of hearing volunteers and average volunteers to test a closed captioning application we will develop. Each volunteer will be given instructions in order to test the usability and performance of the most important functionalities. They will then answer a questionnaire to quantify how easy the application is to use and how readable the display captions are. The questions will be scored from 1 to 5 so that average scores can be generated from the volunteers.

In terms of experimental outcomes and the impact they will have on our hypothesis, the machine learning technique we choose should provide immediate improvement in accuracy over existing methods. This should improve not just the readability of the captions, but also help the grammar and vocabulary, which are the biggest barriers to making captions easy to understand. For the captioning application, the optimal result would be an average score of 4 on the questionnaire. Since the optimal goal of the application is to be able to work will all types of multimedia content, a high average score would prove that the application could see use as a general-purpose captioning tool.