Adrian Marinovich
Springboard Data Science Career Track
Capstone Project #1 Proposal
September 7, 2018

Classification of emotion using the Ryerson RAVDESS video database

What is the problem you want to solve?
The problem is to classify emotions of people in videos using their facial expressions.

Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?
A range of clients may be interested in emotional classification from video of human facial expression. Potential applications allowing for improved decision-making include: market research, to better gauge interest in and reaction to products in order to guide decisions on product design; gaming, to support decisions in creating more immersive and exciting player experiences; and robotics, to allow for more safe and reliable human-robot interactions, so that robots can better infer human intentions and therefore enhance their response decisions.

What data are you using? How will you acquire the data?
The data will be obtained from the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS), which consists of video and audio clips of performances by 24 male and female actors during speech and song in a structured setting, for which validated ground truth emotional expression has been specified.

Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.
Machine learning techniques, including neural networks, will be used to classify 6 'culturally universal' primary emotions (happy, sad, angry, fearful, surprise, disgust), as well as neutral and calm emotions, initially using the video-only, speech portions of the RAVDESS dataset that involve a single spoken phrase in two repetitions, consisting of 384 videos.

The analysis will be limited to classification of still image sequences obtained from the videos, as opposed to true video classification. The classification algorithm will toggle, or trigger, the 6 emotion variables as they reach a certain threshold, and then pick a dominant emotion/rank emotions for the video. A time-series analysis through the image sequence will also be explored. Additional work may extend to analysis of a second spoken phrase, and sung performances of both phrases.

Working within the Amazon Web Services and other GPU-based environments such as Google Colaboratory, the analysis will be conducted using Python tools such as OpenCV, TensorFlow and Keras.

What are your deliverables? Typically, this includes code, a paper, or a slide deck.
The deliverables will include the machine learning code, a brief paper describing the findings of the analysis, and a slide deck, to be made available on GitHub.

Reference:
Livingstone SR, Russo FA (2018) The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic,multimodal set of facial and vocal expressions in North American English. PLoS ONE 13(5):e0196391.
https://doi.org/10.1371/journal.pone.0196391