Machine Learning Engineer Nanodegree

Capstone Proposal

Arun Rajora

May 31st,2017

Proposal: *Facial Expression Recognition using CNN*

Domain Background

It has been easy for humans to recognize faces and expressions on a face. It helps us in communication as it adds more context to a conversation or might convey a quick response to a situation. Adding such capabilities to machines is of great interest as it gives more understanding of the user.

According to a paper by *Hatice Gunes* and *Hayley Hung* (paper can be found on http://www.sciencedirect.com/science/article/pii/S026288561630049X#bbb0005), "Automatic facial expression recognition is a multidisciplinary research field that spans across computer vision, machine learning, neuroscience, psychology and cognitive science. In automatic facial expression recognition research, the most common approach is to classify continuous expressive facial displays according to specific labels, categories or dimensions. Ekman's theory of basic emotions is the most commonly used scheme when creating vision-based systems that attempt to recognize facial expressions of emotions and analyze human affective behavior. The main assumption is that emotions that are felt inside the body are displayed externally via the face, and these in turn can be universally mapped into the six categories of happiness, sadness, surprise, fear, anger and disgust. In reality though felt emotions are not always so visibly manifest because the experience is subjective, nor do they map cleanly to Ekman's six categories."

The same paper by *Hatice Gunes* and *Hayley Hung* (paper can be found on http://www.sciencedirect.com/science/article/pii/S026288561630049X#bbb0005)also states that "In the early 1990s, a number of facial expression recognition researchers had a motivation of revolutionizing the way we interact with technology by enabling it to become more human like. By being able to analyze human emotions through the displayed facial expressions and responding to these in an appropriate and meaningful way, machines would become more intuitive and emotionally and socially intelligent. This paved the way for novel computer vision techniques for analyzing people's facial expressions."

Problem Statement

In this project, the objective is to classify the expression on a human face in an image into the following seven classes- Angry, Disgust, Fear, Happy, Sad, Surprise and Neutral.

We have a dataset consisting of images of people's face which have been labelled with the appropriate expression the face depicts. Our goal is to train a classifier which can classify faces in an image into the expression expressed by that face.

Datasets and Inputs

The dataset has been obtained from a contest on Kaggle. The link for the dataset is https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data .

The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. The task is to categorize each face based on the emotion shown in the facial expression in to one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

The file fer2013.csv contains three columns, "emotion", "pixels" and "Usage". The "emotion" column contains a numeric code ranging from 0 to 6, inclusive, for the emotion that is present in the image. The "pixels" column contains a string surrounded in quotes for each image. The contents of this string is a space-separated pixel values in row major order. The "Usage" column contains one of the following-"Training ", "PublicTest ", "PrivateTest". It depicts whether the data belongs to training, public or private test set.

The training set consists of 28,709 examples. The public test set consists of 3,589 examples. The private test set consists of another 3,589 examples.

According to the contest page, "This dataset was prepared by Pierre-Luc Carrier and Aaron Courville, as part of an ongoing research project. They have graciously provided the workshop organizers with a preliminary version of their dataset to use for this contest."

Solution Statement

For this project, convolutional neural network will be used to classify the face in an image into the expression expressed by the face. Convolutional neural network has been very effective in such kind of problems due to-translation invariance (the model learns the features irrespective of its position in the image) and the model is effective in learning and representing complex abstract concepts.

Benchmark Model

The model will be benchmarked against the following model-

1) A model which randomly classifies faces into various expressions. The CNN should outperform this model with a large margin.

Evaluation Metrics

If $\hat{y}_i$ is the predicted value of the i-th sample and y_i is the corresponding true value and $n_{samples}$ is the number of samples, then the accuracy_score is-

$$\texttt{accuracy}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} 1(\hat{y}_i = y_i)$$

Project Design

The workflow for approaching a solution to the given will be-

- The data will be loaded and some random images and their corresponding label will be visualized.
- The statistics for number of samples for various labels will be analyzed.
- The images will be preprocessed to make each pixel between 0 to 1.
- The data will be split into test/train/validation set.
- A Convolutional Neural Network will be created.
- The CNN will be trained for various hyper-parameters.
- The final accuracy will be recorded for the validation set.
- The final result will be plotted to easily visualize the result.
- Tensorflow and Scikit-Learn will be used for this project.