MACHINE LEARNING ENGINEER NANODEGREE

CAPSTONE PROPOSAL

Namita Atri
August 15, 2018

Domain Background

Facial Expression Recognition is an area of significant research and is proving extremely useful in the field of computer vision and artificial intelligence. Human communication revolves not only around verbal communication but also around non-verbal components, such as reading the emotional state or expression of a person, i.e. happiness, sadness, excitement etc. Therefore, study of FER is gaining momentum as advancement in different aspects of artificial intelligence is taking place, such as human computer interaction, augmented reality, virtual reality etc, that have started attracting attention due to its role in communication. To this end, I propose to implement a solution to Facial Expression Recognition using a deep learning approach.

Problem Statement

The goal of this project is to predict facial expressions to recognize basic human emotions of anger, disgust, fear, happiness, sadness, surprise, neutralilty. I am using the FER2013 dataset downloaded from Kaggle where there are 35777 images divided into training, validation and test sets. I propose to employ a deep learning technique to learn expressions using the above data set and predict emotions of faces taken from new images with reasonable accuracy.

Datasets and Inputs

The dataset used is FER2013 which is taken from Kaggle. This was provided for the ICML challenge, where the image size is 48x48 pixels and is fairly centred around the face to make data appropriate for the task.

There are 35887 rows with 3 columns, the first one indicating the value of the emotion from 0 to 7, second one with pixels of the image, and third signifying its usage: Training, PublicTest and PrivateTest. The first usage category has 28709 images, second and third have 3589 images each. This data should be sufficient to produce acceptable accuracy while training to predict fairly correctly.

Solution Statement

The solution that I am going to employ in this project is Convolutional Neural Networks. CNNs are deep learning algorithms which in the field of computer vision have been given a lot of weightage for the tasks of feature extraction, classification and recognition as they reduce working with models based on physics, and also pre-processing to achieve such models. CNNs employ end-to-end learning directly on the input dataset and are currently being researched to improved results specially in image classification tasks. They work on groups of pixels and therefore are appropriate for this task.

Accuracy is a good method to measure how well the model will perform in comparison with the benchmark model presented next.

Benchmark Model

The benchmark model I am using for this problem is the model provided by Charlie Tang, who is the winner of this Kaggle challenge of facial expression recognition, and achieved an accuracy of 71.161%. I will try to reach an accuracy as high as possible, perhaps in the 50s, as the hardware I am using is quite limited to produce higher results. Charlie Tang used CNN in combination with SVMs, however I plan to use only CNN.

Evaluation Metrics

The accuracy is the ratio of true outcomes, which includes both true positives and true negatives, to the total number of cases examined (true outcomes and false outcomes).

Accuracy (ACC) = (TP + TN) / (TP + TN + FP + FN)

In the context of this project, accuracy can be reframed as the ratio of correct predictions to all predictions.

Accuracy (ACC) = Correct Predictions / All Predictions

The number of correct and total predictions can be easily retrieved while testing on the test data set.

Project Design

I will be using Python 2.7 with Keras, set to Tensorflow in the background.

I plan to incorporate the following steps in the project:

1. Preprocessing of the data set to perform cleaning, reformatting of the images to convert to grayscale and enforce uniformity between all images in terms of size. I'll extract features of the data set, to present the distribution of images within each emotion, with graphical representations of the same. I'll perform one-hot encoding on emotion column as part of preprocessing for the purpose of prediction after the model is trained.

2. I will create a convolutional neural network with layers of Conv2D, MaxPooling, Dense with varying parameters like Dropout coefficient, activation values, padding. I plan to make a model with the idea that the model should contain sufficient parameters such that it should achieve the highest possible accuracy keeping in mind limitations of time, and hardware resources, and the scope of this project.

3. I'll refine the project further by means changing different parameters in the model. This should give better results and will contribute to improved predictions.

4. During steps 2 and 3, I will calculate the accuracy as a means of evaluation metrics to compare the results with that of the benchmark model.

5. At last, I plan to test my model on different images that I will provide, to confirm on visualization how well my model performs.

6. I also plan to support my results with pictorial or graphical representations to better explain the results during implementation.

References

1. A Brief Review of Facial Emotion Recognition Based on Visual Information
2. Challenges in Representation Learning: Facial Expression Recognition Challenge
3. Deep Learning using Linear Support Vector Machines
4. Static facial expression recognition with convolution neural networks