Machine learning engineer nanodegree

Capstone proposal

Facial Expression Recognition

Ola Hamdy Ahmed
November,2019

Domain Background

Facial expression is the facial changes that reflects a person's internal emotional states, intentions and feelings. Humans usually employ different cues to express their emotions, such as facial expressions, hand gestures and voice. Facial expressions represent up to 55% of human-communications while other ways such as oral language are allocated a mere 7% of emotion expression. [1]
Since then, Facial expression recognition (FER) has become a growing area of research with numerous real-world applications such as mental state identification, lie detection, human behaviour understanding and security.
The goal of this project is to apply deep learning techniques to recognize the key seven human emotions: anger, disgust, fear, happiness, sadness, surprise and neutrality.

Figure 1: Example pictures of the seven basic expressions

Problem statement

One of the non-verbal communication methods by which one can understand the emotional state of a person is the expression of his face. As the technology runs our lives these days, the majority of our time is spent in interacting with computers and mobile phones in our daily lives. Therefore, Adding facial expression recognition feature to our devices to expect the user's emotional state can improve human-computer interaction. In this way, it can be used in a healthcare system to detect humans' mental state and improve it by exploring their behaviour patterns.

Datasets and Inputs

In this project, the dataset we will use is Kaggle's Facial Expression Recognition Challenge dataset. This dataset is representative because of its size, unstructured nature of faces and relatively uniform distribution of the data across the seven main human emotions.

The chosen dataset should not only provide a representative number of images, but also should contain data that is uniformly distributed across the race, sex, age and ethnicity, and with a relatively even distribution across the emotions. The Kaggle dataset meets all these aspects.[2]
The dataset attributes:

- 35,887 images
- Image format: 48x48 pixels
- Various individuals across the entire spectrum of:
- ethnicity, age, gender and race, with all these images being taken at various angles.
- The seven key emotions are relatively equally distributed with the one exception being disgust, at ~1.5%.

The dataset contains 2 columns, "emotions" and "pixels". The "emotions" column contains a numeric code from 0 to 6 that represents the emotion of the image (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The "pixels" column contains a string for each image. This string consists of space-separated pixel values in row major order.

To feed the neural network with the inputs we need to reshape them into 48x48x1 numpy arrays, and the emotions value will be converted into categorical values such that each emotion will be represented as 1x7 vector.

Solution statement

The proposed solution is to apply deep learning techniques that have proven to be highly successful in the field of image classification.
I am going to use Convolutional neural network algorithm with the aid of the Keras library. The input of the model will be a human face image and the model predicts the facial expression among the seven basic expressions (anger, disgust, fear, happiness, sadness, surprise and neutrality).

Benchmark model

The benchmark model is Support Vector Machine SVM. SVMs have been widely applied to machine vision fields such as character, handwriting digit and text recognition (Vapnik, 1995; Joachims, 1998), and more recently to satellite image classification (Huang et al, 2002; Mahesh and Mather, 2003) [3].
It is interesting to compare our results with a wide-scoped machine learning algorithm like SVM.
We can expect for CNN to have higher accuracy than the general algorithm. If this assumption is confirmed, we can conclude that, when it comes to image recognition we should give preference to more specialized algorithm in that domain.

Evaluation metrics

The evaluation metrics for this problem is simply the accuracy score. The accuracy is defined by the following equation:

$$accuracy = \frac{tp_1 + tp_2 + tp_3 \ldots tp_i}{n}$$

It is the true positives of all classes over the total number of inputs.

Project design

1) Data preprocessing

I will first explore the data and then apply preprocessing based on the data exploration.
For our dataset, I will need to reshape the input images to meet the requirements of the model. I also can apply augmentations if needed. To the emotions value, it will be converted into categorical values such that each emotion will be represented as 1x7 vector.

2) Data splitting

I will split the dataset into 80% training set, 10% testset and 10% validation set.

3) Model training and evaluation

I will start with a simple model first before training and evaluating it. Then, I will iterate this process trying different architectures and hyper-parameters till reach the most acceptable and reasonable accuracy. Then, compare this model with the benchmark model.

Note: I may add some steps if needed during building the model.

References

[1] “A Review on Automatic Facial Expression Recognition Systems Assisted by Multimodal Sensor Data” https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6514576/
[2] “Kaggle’s Facial Expression Recognition Challenge dataset” https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data
[3] “Classification of Images Using Support Vector Machines” https://arxiv.org/pdf/0709.3967.pdf