Machine Learning Engineer Nanodegree

Capstone Proposal

Luiz Picanço
March 19th, 2019

Proposal

Domain Background

The identification of the emotions associated with facial expressions is an intrinsically human trait, very important in social interactions and communication. Although a task considered easy for humans, it is considered very difficult for software to perform.

Software capable of performing this task satisfactorily has many applications, such as evaluating customer satisfaction in a market research and evaluation of a driver's mental fatigue driving a motor vehicle.

I consider this an interesting problem to solve, being in the field of computer vision, an area that I have a lot of interest.

This project is based on this papers:

- Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution[1]
- Challenges in Representation Learning: A report on three machine learning contests[2]

Problem Statement

The problem to be solved by this project is the identification of which type of emotion is associated with a human face. The types of emotions that can be identified are: neutral, happiness, surprise, sadness, anger, disgust, fear and contempt.

One of the ways to solve the problem is with the use of CNN (convolutional neural network).

Datasets and Inputs

To solve the problem, 2 datasets will be used:

FER2013 - Available on Kaggle, from the "Challenges in Representation Learning: Facial Expression Recognition Challenge" competition [3]. This dataset has 35,888 lines, with the following structure:

- emotion - Emotion numeric code associated with the image(0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). It will not be used in this project.
- pixels - Image data in grayscale with 48x48 pixels.
- usage - Image utilization type(Training, PrivateTest, PublicTest). It will not be used in this project.

A sample of the first lines of the dataset can be seen below:

emotion	pixels	usage
0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121...	Training
0	151 150 147 155 148 133 111 140 170 174 182 15...	Training
2	231 212 156 164 174 138 161 173 182 200 106 38...	Training
4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1...	Training
6	4 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84...	Training

FER+ - Available on Github [4]. This dataset also has 35,888 rows, with the following structure:

- usage - Same meaning of the usage column in FER2013 dataset
- Image name - Image name
- neutral, happiness, surprise, sadness, anger, disgust, fear, contempt, unknown, NF - Number of votes that each emotion received, associating it with the image

A sample of the first lines of the dataset can be seen below:

usage	neutral	happiness	surprise	sadness	anger	disgust	fear	contempt	unknown	NF
Training	4	0	0	1	3	2	0	0	0	0
Training	6	0	1	1	0	0	0	0	2	0
Training	5	0	0	3	1	0	0	0	1	0
Training	4	0	0	4	1	0	0	0	1	0
Training	9	0	0	1	0	0	0	0	0	0

The FER+ is an enhanced version of the FER2013 dataset with 2 main advantages:

- Each image has been labeled by 10 crowd-sourced taggers, providing better quality than the original FER labels
- Since each image has 10 votes, each can have more than one emotion associated with it.

Solution Statement

For the solution of the proposed problem I intend to combine the two datasets, using only the pixels column of the first and the corresponding columns of the emotions of the second dataset. This derived dataset will be splitted into 3 parts: training(80%), validation(10%) and test(10%) sets.

For the construction of the prediction model I will use Deep Learning through the use of CNN (convolutional neural network). CNN is a type of neural network widely used in solving image classification problems.

To the first layer of the network I will create an input with a shape of 48x48x1, relative to the amount of pixels of the image and the grayscale. In the last layer I will create an output with 10 nodes, with each one representing a different emotion. With softmax as a activation function, we can have more than one output on the network, associated with a probability.

Another approach can also be used, associating each facial image with the emotion with the majority of the votes.

Benchmark Model

Given the image of a human face, the model must predict a human emotion associated with that face. Since 30% of the dataset is associated with a neutral emotion, I will perform a naive approach that all the images in the dataset are related to that emotion.

The results of the model will be compared with the naive predictor to evaluate the performance.

Evaluation Metrics

For the evaluation metrics I will use a confusion matrix to identify the false positives, false negatives, true positives and true negatives. F1 score will also be used to measure the performance of the model[5].

```
Precision = TP/TP+FP
Recall = TP/TP+FN
F1 Score = 2*(Recall * Precision) / (Recall + Precision)
```

Project Design

For approaching my solution I intend to use Python 3, Keras 2.2.x and Pandas, following this steps:

1. Read the 2 datasets
2. Perform an EDA on the two datasets
3. Join the first dataset with the second, creating a unique dataset, as stated in the "Datasets and Inputs" section.
4. Perform an EDA on the new dataset
5. Separate the dataset into 3 parts: training, evaluation and testing
6. Build a CNN model using Keras
7. Use the training and evaluation sets to train the model.
8. Test the network with the test set
9. Perform the evaluation metrics and compare with the benchmark model, using the test set

References

[1] Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution[1](https://arxiv.org/abs/1608.01041)

[2] Challenges in Representation Learning: A report on three machine learning contests[2](https://arxiv.org/abs/1307.0414)

[3] Challenges in Representation Learning: Facial Expression Recognition Challenge (https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data)

[4] FER+(https://github.com/Microsoft/FERPlus)

[5] An introduction to ROC analysis(https://people.inf.elte.hu/kiss/11dwhdm/roc.pdf)