Udacity Machine Learning Engineer Nanodegree Program

Capstone Proposal

Xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Wong Chee Howe

Oct 2021

1. Domain Background

With the pandemic, the default mode of communication has been via video conference and remote meeting. This mode of communication has allowed for it to engage wider folk but has it been able to engage deeper?

Facial expressions are both a natural and direct means for human beings to convey their emotions and intentions. It is used to express non-verbal communication to one another whether deliberate or unintentionally. According to experts, these nonverbal signals make up a huge part of essential communication[1]. From our facial expressions, the things we *don't* say can still convey volumes of information[2]. Understanding and interpreting emotions among social interactions has became an important skillsets

Researchers are particularly interested in developing techniques to interpret, code facial expressions and extract these features in order to have a better prediction. This has widespread applications and implications for fields in retail, surveys and even negotiation. Even more so in this pandemic era, where meetings, presentations and interviews are done online and information of intention can be gathered by buyers and sellers alike.

.

2. Problem Statement

With the remarkable success of machine learning and particularly deep learning, the different types of architectures of this technique are exploited to achieve a better performance. The problem is a computer vision supervised learning classifier problem based upon multiple classes. From an image, one must predict the correct classification of the emotion display (of which there are originally 7). The emotion state display in this image are: (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). We will have to build a model to correctly assess the emotion display by the faces expression in the picture

This project has twin purpose

1. It is to make a study automatic facial emotion recognition FER via deep learning
2. To fulfill the requirements of Udacity Machine Learning Nanodegree

3. Dataset and Input

The dataset is downloaded from Kaggle, based on a competition: “**Challenges in Representation Learning: Facial Expression Recognition Challenge”**

From the description, we can pick out the following information

- Image pixels is about 48x48 gray scale image
    - Little to no need to normalize or scale the image
    - We will have to convert the string of pixels values into images
- Faces are pre-registered with faces occupying more or less central and occupies the about the same amount of space
    - Little or no image preprocessing would be needed, eg performing face detection and then centralized them
- The training and test database contains 2 columns, one would state the emotion displayed and the other would be a string of pixel values
    - From the brief look at the dataset, we can observe that there is no over dominating class with a large number of images in its dataset, though smile is almost twice the number of the next biggest class
    - The emotion state of disgust is very little. This might mean it might not form a representative proportion in the dataset and class imbalance is a serious problem we would have to consider for this data set. We would need to either augment it or perhaps to think of ways to discard some of the images in other sets
    - Class weights [5] may also be added to correct the imbalance
    - Some processing might be necessary to make them as image in order to better visualize them

4. Solution Statement

Convolutional Neural Networks (CNN) can be used to make a model. CNN is a part of deep neural networks and is great for analyzing images.

We will could adopt the following to improve the result compared to the techniques known then in the competition

1. Using a different architecture like ResNet
2. Using and tuning different learning rate and optimizer
3. Using various augmentation techniques to improve situation of class imbalance

5. Benchmark Mode

This is a rather tough data set when it was first introduced, human accuracy on a subset of the dataset was 65±5%[4]. And the final winner of the competition has achieved a result of 71%. For a novice we aimed to have a baseline goal of 60% and performance goal of around than 65%

6. Evaluation Metrics

We will try to use accuracy as with the original competition as the original competition does. However with there exist imbalances with the dataset, we will probably need to include precision and recall, and most probably F1 score in our studies

7. Project Design

We will divide the project into the following phases:

- Data Gathering & Preprocessing
- Training Sets
- Modelling and Training
- Testing and Evaluation
- Wider Testing

Data Gathering & Preprocessing

After downloading the database from the website, we will convert the pixel values into images for better visualization and checking. We will then band the training images into individual class folders (about 7 of them). After which we will randomly sampled each class to produce representational image for the class. This is done to ensure the previous step is performed in correct manner

Training/Test Sets

First we will start off by defining our mini batch and it's batch size. Next we will have to perform data augmentation in order to handle the problem of data imbalance. Here are some methods of augmentation we will likely choose from

1. Horizontal flip
2. Random brightness
3. Horizontal/Vertical shift
4. Random Rotation

Of which, the last two might be done in minute manner or not at all due to the small pixel size of images

Class weights might be used based on circumstances

Modelling

We will create the model based on various CNN architecture and have fully connected layer at the end, with soft max activation function to give us the probabilities value of the 7 classes and the max probabilities will be taken for the class

We will probably first try to use Adam as adaptive learning rate *optimization* algorithm and categorical_cross entropy as loss function

Testing

We will use the test set to evaluate our model.

Deployment Testing

We will continue to test our model to verify their usefulness. The wild images could come from

- Using open-sourced images in internet (perhaps a few of them)
- Use video cam to capture people facial expression (albeit with their consent)

8. Reference

1. Tipper CM, Signorini G, Grafton ST. Body language in the brain: constructing meaning from expressive movement. *Front Hum Neurosci*. 2015;9:450. doi:10.3389/fnhum.2015.00450
2. Foley GN, Gentile JP. Nonverbal communication in psychotherapy. *Psychiatry (Edgmont)*. 2010;7(6):38-44.
3. https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data
4. Challenges in Representation Learning: A report on three machine learning contest. ,http://arxiv.org/abs/1307.0414
5. https://towardsdatascience.com/address-class-imbalance-easily-with-pytorch-e2d4fa208627