Facial Emotion Recognition

By Anju Pandey

What is Facial Emotion / Expression

We, as a human being express ourselves in many ways. Verbally and with nonverbal behavior (for example our facial expression, body language).

Professor Albert Mehrabian, in 1967, formulated the 7-38-55% communication rule. This rule says that only 7% of feelings we communicate through the words we use, 38%through tone of the voice and 55% through our body language.

Emotional AI Technology:

What is it? In simple words, it is machine or artificial intelligence which can comprehend human emotions.

Understanding human emotions/feelings is a very complex function even for human being so just by reading facial expression, comprehending the emotion is a big challenge. Because expressions can be faked, true emotions can be suppressed.

So, depending upon the area or our problem statement, Emotional AI can be helpful or just not useful at all.

Success Stories of Emotional AI technology

There are many areas where Emotional AI already has been implemented and it is providing good result. This technology can help businesses capture people's emotional reactions in real time. Below are some examples where Emotion Detection and Recognition (EDR) system is being used:

1. **Subway ads**- Brazil's Yellow Line of the Sao Paulo Metro deployed **AdMobilize emotion AI analytics technology** to optimize their subway interactive ads according to people's emotions.

2. **Travel recommendation**-Skyscanner uses **Sightcorp's emotion AI technology** to detect and measure facial expressions like happiness, sadness, disgust, surprise, anger, and fear.

3. **Blood pressure detection**-**The American Heart Association** developed an app using **NuraLogix emotion AI algorithms** to detect blood pressure levels from 2-minute videos.

EDR is evolving and market researchers are indicating that in coming years it will have rapid growth.

Capstone Project (Facial Emotion Recognition)

With this capstone project (Facial Emotion Recognition), goal is to build a model which can distinguish between different facial expressions.

Dataset:

Provided dataset has four categories:Happy,Sad,Surprise and Neutral. Data is divided into three parts: Training, Validation and Test. Total 20,214 images are provided with 74.74 % , 24.62% and 0.633% image distribution among Training, validation and test respectively.

Training data contains following number of images under each label/category:

- Happy : 3976
- Sad : 3982
- Neutral : 3978
- Surprise : 3173

Images under ‘surprise’ is less than the other categories. But, we can say that data is equally distributed as difference is not very high. Other categories are having approx. 26% of data and ‘Surprise’ is having 21% of data.

Model Architecture:

ANN vs. CNN:

If we have to solve image classification problem with **ANN then below are the drawbacks:**

- First, we need to convert 2-dimensional image into a 1-dimensional vector. And this process creates **huge number of parameters** to be trained.

- In images, pixels are arranged in certain way and that special arrangement is called Spatial features. When image is converted into 1-dimensional vector, **ANN loses the spatial feature**.

- We do not need whole image or all the pixels of image for classification. ANN lacks the functionality of extracting only the relevant features and thus we get huge number of parameters which are not needed.

CNN uses filters(kernels). Convolution or filter is applied to input image and relevant features are extracted. It captures the spatial features in relation to other objects in the image. CNN's 'maxpooling' layer reduces the image dimension but keeps the more prominent/ important features.

As, this project is about image classification, we will use CNN. Images can be colored or grayscale. Let's analyze which option is better for this project.

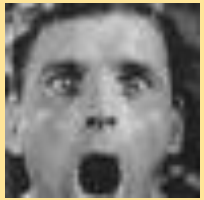
RGB vs. Grayscale image

RGB or colored images are having 3 channels. i.e. Red, Green and Blue whereas Grayscale images have only one channel. Let say, we have color image of 28*28 pixels then in CNN, parameters for this image would be (28*28*3) whereas for grayscale image it would be (28*28*1). For human expression or emotion classification problem, we do not need colored images. **Grayscale images can provide sufficient information for classification.**

Now, **question is how machine is going to identify the expressions** using CNN and grayscale image? Answer is **“Important Features”.**

We say someone is happy or sad or emoting some other feeling by looking at their faces. But what does the face say? What do we observe? Look at the below pictures of different expressions.



We can see that when someone smiles, lip corner goes up, eyes squinch. In case of sad expression, lip corner goes down. In case of surprise, mouth is opened with wide eyes while in neutral expression lip stays straight, eyes are normal, no stressed or wrinkled forehead.

So, these are our features(and many more. Combination of features decides one expression) and we want CNN to extract these features and classify the images.

CNN Models:

In this project, we have built three models. Below table represents their accuracy level for training, validation and test dataset.

If we observe the accuracy level for test data then model3 is the highest performer. However, 70% accuracy is not a good percentage.

In project file, predictions were done on many images and results were clear indicative that models were not able to extract all the important features to distinguish between the expressions.

	Train	Validation	Test
Model1	62.48	64.72	61.72
Model2	83.31	67.81	69.53
Model3	68.42	69.54	70.31

Two possible reasons could be:

1. **Model Architecture**: When we build a model then there is **no said rule** for filter size, number of layers, types of layers and all other parameters and hyperparameters involved. We architect, test and tune it until it produces desired result.

2. **Insufficient training data**: we need lot of data to train a model. Let say, our model never saw a image where person is happy but eyes are closed or covered or model encountered such data in very little numbers. So, when model has to predict such image then it is going to commit mistake.

Conclusion: Even this amount of data can give better result than what we are getting because while predicting sometimes model 1 was committing mistake and sometimes model 2. Hence, We can say that both are using some common and some completely different features for classification. So, certainly model who can extract all these features can give better result. Therefore, before flooding dataset with more data, model architecture should be redesigned.

References:

https://www.bl.uk/people/albert-mehrabian

https://builders.intel.com/ai/membership/sightcorp

https://www.admobilize.com/

https://research.aimultiple.com/emotional-ai-examples/

Thank You !