Emotion Detection

Şeyma YILMAZ [1]
Mücahit FINDIK [2]
Yusuf Emre GENÇ [3]

Fundamentals of Machine Learning Lab.
Department of Computer Engineering
Hacettepe University, Ankara, TURKEY

1. Abstract

In this paper, we mention about the project we are developing to classify the basic emotions of people according to pictures. Our data set consists of 37k images. 30k of this data is used for training and the rest is used for validation. The size of the images is 48x48. When developing our model we use the CNN algorithm, which is very suitable for classifying images. We will use the confusion matrix to evaluate model accuracy.

2. Introduction

1. Introduction People express their thoughts and emotions not only by sound but also by facial expressions. Facial expressions are very important for clearly expressing how we feel to people and to clearly understand what they are feeling. From the moment people were born, they began to express their troubles with facial expressions. Then, they continued to learn to express their emotions and understand the emotions of people by using facial expressions throughout their lives. In doing all that, people use their intuitions and observations. In this project, we aim the machines to perceive the emotions that people perceive with their intuition and observations. One of the main contributions of this project is to provide infrastructure for many important projects by classifying people's emotions according to their pictures. For example, when conducting psychological analyzes, facial expressions of people are of great importance. Using a machine-learning algorithm to determine the emotion of a person in a picture makes these investigations independent of the intuition of the researcher. Thus, more objective results are achieved in a faster research process. Another example is that an assistant who detects moments when people are unhappy. It can plays music for them or opens entertaining videos and so it keeps people away from depressed feelings. Emotion classification can be used in many imaginable projects like these. Another contribution of the project is to transform the subjective comments perceived by the intuition of the people into objective interpretations with the machine learning algorithm, as in the example mentioned above. Besides, even in an environment where people are not present, the emotions of people can be perceived, that is, the emotion perception mechanism can be free from the people.In other parts of the page, we will describe in more detail below topics:

• In section 2, other important works similar to our project topic,

• In section 3, the machine learning method we will use,

• In section 4, the data set we will use for classification, and how to evaluate the results.

3. Methodology To Be Employed

We will make our project using CNN. In general, when we examine similar projects to our project, CNN structure was used to classify the photographs. When we look at the related projects, we understand how important dataset is. Therefore, if there is a problem with our dataset, we plan to move to a different dataset. CNN algorithm consists of Convolution Layer, Pooling Layer, Fully connected layer. We will briefly talk about these issues.

Convolution layer:

The role of this structure in CNN is to facilitate the process-

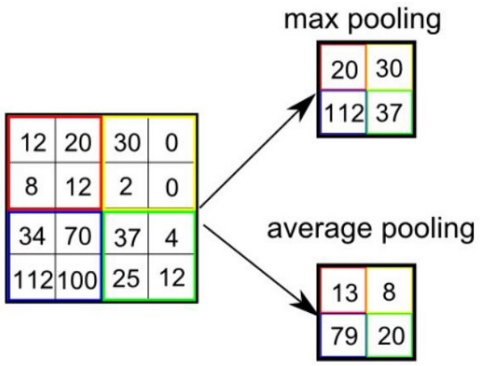

Figure 1. A sample figure. Max-Average Pooling (taken from https://towardsdatascience.com).

ing of the photograph without losing the features that are important for better results when classifying photos.

For example, the photos we have are 48 * 48. We need to create a total of 48 * 48 = 2304 neurons to indicate the importance of each pixel for these photographs, and we need to determine 2304 weight for them. For back-propagation we need to update 2304 weight in one layer. This process is very laborious and forces your computer hardware as functionality.

CNN uses a matrix called the kernel to solve the problem described above. Let's assume that the kernel is 5 * 5 for example. It takes important features of the photo by hovering this kernel in the photo. 25 weight is specified for each kernel. This reduces the size of our photo and assimilates the photo for subsequent layers. This also gives us functionality.

Pooling Layer:

This layer is used to reduce the size of image matrices from Convolution Layer. As a result of this diminution, the computational function required to process the data is reduced. This layer also removes the dominant features of our photograph. Thus, it is very effective in training our model.

For pooling layer, max pooling and average pooling are used. To put it briefly, max pooling returns the largest value in the area covered by the kernel in the photograph. Average pooling returns the average of the values the area covered by the kernel in the photo.

Fully Connected Layer:

Neurons in one layer bind to each neuron in the next layer. Our photo is smoothed after passing through the layers described above and this layer is used to classify the images.

We think that we will use some data augmentation methods. Because our disgust class photos are quite few (there are 437 photos in the train dataset, 111 photos in the validation dataset.) So we think that our model will be less accurate in that class. For this reason, we intend to increase these photos. To do this, we can increase the number of photos by rotating photos of this class or by cutting certain areas.

We are planning to create our own CNN architecture in our project. When we examine the projects related to our projects, we see that 6 to 10 CNN layers are used. We consider this number of layers according to our validation accuracy and apply the most appropriate one for our model.

As different CNN architectures, VGG16, RESNET50, AlexNet can be used.

4. Related Work

Detection emotions from facial expressions of people is a popular area of machine learning. So that there are many related works with our project. We will share five of them such as [1, 2, 3, 4, 5]. In project [1], three modules are used. The first module uses Markov random fields (abbreviated as MRF). By using MRF, model detects the skin and segments images. The second module finds the location of eyes and mouth. In the last module, emotion is detected by using edge detection and the location data of eyes and mouth. In [2], there is another important way ,as well image, to detect emotion that is auditory data. This idea makes the project multi-model. This hybrid model is trained by video recordings. In [3], In this article, the usage areas of the model that we want to develop are given. One of them showed during the experiment that the emotional scenes discovered by the films had a greater impact on the audience. Another area of usage is that the speaker changes the flow of the speech by taking into account the emotion that the listener is getting from facial expressions. Emotion is detected through facial feature recognition in [4]. Viola-Jones and Harris algorithms are used. Viola-Jones helps to detect faces in way that's so accurate and fast [5]. Harris algorithm is used in lots of computer vision problems. It helps to extract the edges from image. This algorithm is used to detect the edge of face which caters to calculate slopes between eyes and mouth.

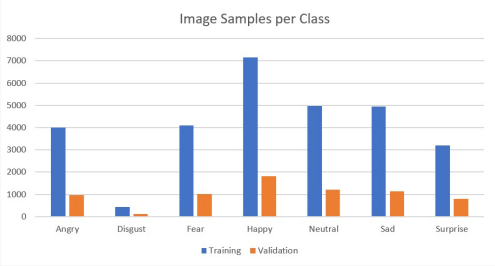

Figure 2. A sample figure. Data Set Distribution .

		Predicted Number						
		Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Actual Number	Angry	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}
	Disgust	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}
	Fear	x_{31}	x_{32}	x_{33}	x_{34}	x_{35}	x_{36}	x_{37}
	Happy	x_{41}	x_{42}	x_{43}	x_{44}	x_{45}	x_{46}	x_{47}
	Neutral	x_{51}	x_{52}	x_{53}	x_{54}	x_{55}	x_{56}	x_{57}
	Sad	x_{61}	x_{62}	x_{63}	x_{64}	x_{65}	x_{66}	x_{67}
	Surprise	x_{71}	x_{72}	x_{73}	x_{74}	x_{75}	x_{76}	x_{77}

Figure-2

Figure 3. A sample figure. Predict-Actual Class .

5. Experimental Evaluation

We will use the data set of 37k images that we reach through kaggle. 30k of the data set is reserved for training and the rest is reserved for validation. The classes in the data set consist of people's most evident emotions which are angry, disgust, fear, happy, neutral, sad and surprise. Given the number of images and the number of labels, we think that the size of the data set is sufficient for training. In Figure-1 you can see the data distribution for each class. As can be seen from the figure, the data is not distributed properly. There is very little data, especially in the disgust class. This may degrade the performance of the model. We plan to overcome this problem by working on hyperparameters such as the number of layers of CNN and using data augmentation.

We will use the confusion matrix to evaluate our results. A confusion matrix is a measurement tool that provides information about the accuracy of predictions and hence the success of our model. The reason why we use this evaluation method is that the complexity matrix provides easy-to-understand information about the accuracy of the measurement. The Generalized Confusion Matrix for Multiple Classes determined by the number of classes in our data set is as follows.

According to the information we have obtained from the reference specified, the general accuracy is calculated according to the following formula. However, we will evaluate the model we created with the confusion-matrix method in the sklearn.metrics module. Overall Accuracy = (x11 + x22 + x33 + x44 + x55 + x66 + x77) / Total Number of Testing Entries [10]

References

[1] http://www.bmva.org/bmvc/1988/avc-88-023.pdf

[2] https://ieeexplore.ieee.org/abstract/document/647126

[3] https://onlinelibrary.wiley.com/doi/abs/10.1002/

[4] https://web.stanford.edu/class/ee368/Project$_{A}utumn_{1}617/Reports/report_{p}ao.pdf$

[5] https://www.cs.cmu.edu/ efros/courses/LBMV07/Papers/viola-cvpr-01.pdf

[6] https://www.researchgate.net/publication/310799885$_{G}eneralized_{C}onfusion_{M}atrix_{f}or_{M}ultiple_{C}lasses$

[7] https://en.wikipedia.org/wiki/Convolutional$_{n}eural_{n}etwork$

[8] https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-

[9] https://mc.ai/veri-artirma-data-augmentation/

[10] https://www.researchgate.net/publication/310799885$_{G}eneralized_{C}onfusion_{M}atrix_{f}or_{M}ultiple_{C}lasses$

[11] https://towardsdatascience.com/classifying-facial-emotions-via-machine-learning-5aac111932

[12] https://medium.com/@hinasharma19se/facial-expressions-recognition-b022318d842a