MACHINE LEARNING ENGINEER NANODEGREE

Capstone Proposal

Mariam Ashraf

July 31st, 2019

Proposal

DOMAIN BACKGROUND

This project is for Facial Expression Analysis as part of the Kaggle competition Challenges in Representation Learning: Facial Expression Recognition Challenge from ICML 2013[1], later published in the Neural Information Processing journal.[2] The aim of the project is to construct a model that will be able to infer how a person is feeling from a photo of their face. This application can be helpful when implemented in robot-human interactions so the robot can have more information input to use when deciding on what to say or offer to the person speaking to them. This is usually done by defining emotions in a few categories, instead of the range humans are more used to, such as anger, happiness, and so on.

PROBLEM STATEMENT

The problem to be solved is to categorize a photo of a person's face into one of 7 emotions: Anger, Disgust, Fear, Happiness, Sadness, Surprise, and Neutral. The model should be able to make this decision for grayscale photos with a clear front shot of a person's face centered in the square photo.

DATASETS AND INPUTS

The dataset used is provided by the competition. It consists of 48x48 square grayscale photo distributed as follows, 28709 photos in a training set, 3589 photos in a public test set, and a private test set of another 3589 photos that was used by the competition to determine the winner. The input is a csv file with the first column signifying the emotion in the photo via a number between 0-6 inclusive, the numbers are in the order stated in the problem statement. The second column has the grayscale value of each pixel separated by spaces. So each photo has a vector-like structure of 2304 numbers each signifying how white or how black each pixel is. The last column has either "Training", "PublicTest, or "PrivateTest" to show which group the photo belongs to.

SOLUTION STATEMENT

There are several possible ways to design a model to solve this problem with the given data. First, it is in my opinion that a Neural Network is a the most befitting candidate for a classification problem such as this since it is able to discern features in the data provided with

increasing complexity and should be able to quickly separate the 7 emotions stated. A second potential model is a decision tree or forest, however, it is not as recommended since it is more prone to overfitting and is not as versatile as a neural network.

A Neural Network can either be an MLP or a CNN. An MLP will need little adjustments to the input data since the photos are already flattened. Another option is to use a pre-trained network such as Resnet-50 and only make minor changes to avoid the extensive training time. The chosen solution so far is a newly designed MLP even if it will take more time to train.

BENCHMARK MODEL

In the competition, the winning teams had accuracies between 65% and 71%. In the paper published later, it was stated that human accuracy on the same set was around 65%. Since humans are especially adept at this task, a model that reaches a similar accuracy to humans should be acceptable. Furthermore, a null model garnered results of about 60%, and an ensemble of null models reached 65%. Since an untrained model reached 60%, a fully trained model should achieve higher.

EVALUATION METRICS

As per the competition, the evaluation metric to be used will be accuracy. Since all the data provided is labeled, it should be easy to check how many times the model was correct. Further analysis can be made to each emotion to find out if the model has any weakness, for example if the model mistakes surprise for happiness frequently. This might be a telling sign that the model is could do better with slight altering, or it may point out a shortcoming in the dataset used for training.

PROJECT DESIGN

As per the comments made in the Solution Statement section, my preferred models are the MLP, the CNN, and the retrained network; it may be possible to test all three methods to decide which yields the best results. The MLP will be the easiest to start with since the data is already flattened and will need only be imported before feeding it into the network. I also prefer a new design to transfer learning, so next to be tested is the CNN. This will require the data to be re-set into a matrix format. There is possibility a CNN receives a higher accuracy scoring since it will take into account the neighboring pixels when extracting features, not just a line of the pixels like in the MLP. Thirdly, transfer learning may be given a try if sufficient time is left.

REFERENCES

[1] "Challenges in Representation Learning: Facial Expression Recognition Challenge | Kaggle", *Kaggle.com*, 2019. [Online]. Available: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/overview.

[2] Goodfellow, I., Erhan, D., Carrier, P. and Courville, A. (2013). Challenges in Representation Learning: A Report on Three Machine Learning Contests. *Neural Information Processing*, pp.117-124.