Machine Learning Engineer Nanodegree

Capstone Proposal

My Nguyen
August 3rd, 2017

Proposal: Detecting Emotions through Facial Expression

Domain Background

Recent advancements in the field of computer vision, particularly image recognition, have given new waves of progress to the task of emotion detection through facial expressions. Convolutional Neural Networks (CNNs) have been in the center stage of these improvements, outperforming other methods. Notable examples include the Emotion in The Wild (EmotiW) challenges, where top submissions have consistently employed CNNs [1],[2],[3], and the Facial Expression Recognition (FER) competition from Kaggle, where the top three submissions all used CNNs [4].

This project is inspired by AutoTutor, an Intelligent Tutoring System (ITS) that is both cognitively and emotionally intelligent. It can carry conversations with students, know when the student is frustrated or bored and adjust its teaching methods accordingly [5]. It is my hope that CNNs can help the tutor better recognize emotions. Although learning emotions are not the same as the basic emotions represented in this Kaggle competition [5], due to the lack of labelled datasets of learning emotions, the FER Kaggle dataset is a good starting point.

Problem Statement

The task at hand is to classify seven basic emotions based on facial expressions: angry, disgust, happy, sad, fear, surprise, and neutral.

Datasets and Inputs

The dataset was highly curated by human labelers from a pool of google images. Faces were cropped out. Images were resized to 48x48 pixels, and converted to grayscale. There are 35887 images in the final dataset, with 4953 of "Anger", 547 of "Disgust", 5121 of "Fear", 8989 of "Happiness", 6077 of "Sadness", 4002 of "Surprise", and 6198 "Neutral". Further details about the dataset can be found here [6].
This dataset has the advantage of being spontaneous and realistic as opposed to posed and perfect lighting conditions as in laboratory settings. One disadvantage could be that the perceived emotions are not the same as expressed emotions.

Solution Statement

Use CNNs to train the classifier with the 1st place solution as a starting point [7].

Benchmark Model

I hope to be among the top 10 of the leaderboard, ie. minimum of 62% accuracy.

Evaluation Metrics

Accuracy rate

Project Design

The project follows this basic structure:
- Read in data from csv file.
- Split into training, validation, and test sets. As the whole dataset was released, I used the public leaderboard set as the validation set, and private leaderboard as the test set.
- Format data as necessary: into numpy arrays, one hot encode, remove bad data, etc.
- Perform data augmentation and preprocessing: center and scale, mirror images, rotation, shifting, etc.
- Specify a CNN structure with the 1st place solution as a start.
- Try different parameters of learning rate decay, optimization engine, regularization.
- Use Tensorboard for visualization and debugging.
- Use an ensemble for testing.

References

[1] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, ICMI '15, (New York, NY, USA), pp. 435–442, ACM, 2015.

[2] B. Kim, J. Roh, S. Dong, and S. Lee, "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition," Journal on Multimodal User Interfaces, pp. 1–17, 2016.

[3] Y. Fan, X. Lu, D. Li, and Y. Liu, "Video-based Emotion Recognition Using CNN-RNN and C3D Hybrid Networks," Proceedings of the 18th ACM International Conference on Multimodal Interaction, pp. 445-450, ACM, 2016.

[4] https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge

[5] D'Mello, S. K., Graesser, A. C.. 2012. AutoTutor and Affective AutoTutor: Learning by Talking with a Cognitively and Emotionally Intelligent Computer that Talks Back. ACM Trans. Interactive Intelligent. Syst. X, X, Article XX (Month 2011), X pages.

[6] [1307.0414] Challenges in Representation Learning: A report on three machine learning contests https://arxiv.org/abs/1307.0414

[7] https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/discussion/4676#25019