Proposal: Deep Learning Neural Network to Detect Deception in Video
Ledger West

Problem

This project will seek to detect when a human subject is lying or telling the truth based on video input without audio. The best current method for identifying human deception is a polygraph, but it requires physical contacts and can produce inconclusive results due to a variety of stress or observation related factors. Humans as lie detectors are pretty bad. In a recent (limited) study, humans correctly identified human deception about 50% of the time (i.e. no better than randomly guessing lying or not lying). This test was across full video, audio-only, video-only, and transcript, and detection performance was about the same across all mediums.

A research group at University of Michigan recently developed a multimodal decision tree and random forest model that achieved 75% accuracy using video, audio, and thermal inputs. The video-only model achieved 68% accuracy.

Some specific questions this project will explore include:

1. Can we train a deep neural network to outperform the University of Michigan lab's decision tree and random forest model on video-only detection?
2. If the neural net performs well, is it also more useful, given that the UMich lab's model required 30+ one-hot encoded features derived from the video. In other words, will the neural net predict faster?

Potential Client

There a*re four potential clients for this project.*

1. Media and Journalists
2. Law Enforcement
3. Intelligence Community
4. Lawyers

Data

I plan to use two primary sources of data.

1. Real-Life Deception video dataset: This is the Michigan Lab's dataset of 121 labeled video clips. Average clip duration is 28 seconds.
2. Imagenet: This is a dataset of nearly 15 million labeled images. I won't directly use Imagenet, but will be using the weights learned from InceptionV3's training on Imagenet to initialize the weights in my neural net.

Approach

Based on some initial research, the working approach at this point includes:

1. Video input preprocessing
    a. Down sampling of frames
    b. Standardize frame dimensions by cropping or down sampling resolution

---

2. Build neural net using Keras on Tensorflow
3. Transferring learning from the InceptionV3 CNN trained on Imagenet
4. CNN layer for each frame >> LSTM layer >> Fully Connected Layers >> Sigmoid
5. Tensorboard for NN architecture inspection
6. Build and train model with cloud GPUs on FloydHub
7. Publish report using Distill publication framework

Deliverables

The deliverables for this project will be:

1. A report published using the Distill web publication framework
2. Code for a neural net that detects human deception in video
3. Simple web demo of other videos in the wild
4. If time permits, interface to run lie detector on user-supplied video clips (like YouTube video clips)