InfoGAN for Unsupervised Facial Recognition

Kastan Day, Tim Nguyen

November 2017

1 Proposed Learning Task

We aim to classify images of people's faces in an unsupervised learning environment. The algorithm will be given a training data-set of face images, possibly taken from a robot's on-board camera, and will attempt to group those photos into distinct categories to represent distinct people. We expect that given a categorical variable with which to learn these representations, InfoGAN will learn distinct representations of each person.

Our unsupervised learning environment also allows us to specify continuous variables from which we can learn continuous features, like hair color, or face shape or even height. We find it much more difficult to predict precisely what the network may learn with continuous representations of the data. Successfully learning to recognize human faces from unsupervised training would be a fascinating result, and a useful product for industries of all types.

2 The Adaptive Method

In order to accomplish our learning task, we will use an extension of Generative Adversarial Networks called InfoGAN. GAN's are a class of AI algorithms intended for unsupervised learning which consist of two neural networks battling in a co-evolutionary arms-race. The two networks consist of a discriminator and a generator. The generator continually generates images in order to trick the discriminator into believing it's a real image. The discriminatory, on the other hand, is responsible for distinguishing between real and generated images.

A natural extension of GAN, is for the network to learn a *conditional* generative distribution. This variant is called CGAN. InfoGAN, however, extends CGAN by providing not only a helpful conditional variable, but the system learns this conditional automatically.

3 Hypothesis and Outcome

We hypothesize that based on the output of InfoGAN, this network would successfully be able to distinguish between different faces and perform a different type of facial recognition. InfoGAN will be able to draw out *disentangled representations* from the training set of facial images and perform a hybrid of facial recognition and feature distinction. For example, the neural networks should be able to output differences in faces, such as hair color and eyes, face shapes, widths, lengths, or any distinctive feature.

4 Proposed Experiments

We propose an experiment to determine if InfoGAN is capable of creating disentangled representations of human faces, given a training set of faces and a categorical variable into which it can fit distinct representations.

Given a categorical variable with 10 elements, and a training set that contains photos of 10 distinct faces, we expect InfoGAN to fit a representation of each face to each element.

If unsupervised facial recognition were possible, it could be applied to many fields such as by enabling businesses to recognize customers and suggest purchases, classrooms that can recognize students and take attendance and much more.

In this experiment, the purpose is clear, but the best path to achieving it is less so. Below we articulate several important variables we will explore in creating the most robust unsupervised face detection system, which we hope may also be useful in extracting other relevant features.

Key variables:

1. The nature of the training images
   - We will use self-gathered images from a camera looking at faces in natural human environments, extracting images from the video feed every few seconds.
   - We will also test images from online datasets of faces, yet this comes at the cost of a more consistent learning environment around the person's face or regarding the camera used to take the photo.

2. The size of the categorical variable (which represents the number of people to recognize)
   - We hypothesize that using a smaller categorical variable with fewer faces to categorize will yield more accurate results as the probability of two similar faces appearing in a dataset increases with the size of the dataset.

3. The size of the categorical variable relative to number of unique faces.
   - We would find it interesting to test what the categorical variable would learn if its size does not match the number of unique faces it is shown. It could still learn faces, but fewer or potentially with multiple similar faces mapped to each variable element. It could also learn completely different features like the hairstyle of the person it sees.

4. Multiple categorical variables
   - Possibly if 10 faces are present in a dataset that two categorical variables of 10 faces should be learned, then averaged to produce the best results.
   - Multiple categorical variables might also learn features like hairstyle or shirt color.

5. The value of continuous representations
   - We hypothesize that continuous variables might yield very useful features that could correspond with a person's face like their hair color, eye shape, face shape, nose size, or how much someone is smiling. - These features could be very relevant to any real-world application of unsupervised facial recognition.

6. Boolean representations
   - We hypothesize that things like gender or the property of 'wearing glasses' could become represented in the Boolean variable, but these results are not obvious to predict.

5   Experimental Analysis

During the process of training our network, InfoGAN's discriminator net will output probabilities that a given image would is from the real dataset. Additionally, the generator network continually creates images with increasing realism over the duration of training. These generated images, combined with the statistical output of the discriminator, is the foundation of our plan to prove our hypothesis.

6   Additional Experiments

Inspired by our recent classwork on style transfer, as a secondary project we are also deeply curious as to what InfoGAN could learn about 2d artwork, and particularly paintings. InfoGAN's unsupervised learning

combined with its generative capabilities make its application discovering and reproducing artistic traits, or mixtures of them, particularly compelling.

In an experiment we might expect InfoGAN to use continuous variables to learn continuous features like painting texture (brushstroke depth), intensity of color, or even how angular vs. rounded a painting is. We might also expect that InfoGAN could use categorical variables to learn different painting styles, or artists, when given matching dataset. Yet what might we expect Boolean variables, or very large categorical variables to learn?

Finally, we could then train InfoGAN on various styles of art and ask it to adjust various parameters outside of it's training set. For instance, you could adjust the continuous variable of brushstroke size on a painting and watch its texture expand and recede. That would be an absolutely fascinating phenomenon to observe.

References

[1] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. *CoRR*, abs/1606.03657, 2016.