InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Netss

published:2016-06

authors: UC Berkeley, OpenAI

> A **disentangled representation**, one which explicitly represents the salient attrubutes of a data instance, should be helpful for the relevant but unknown tasks.

> A **disentangled representation** can be useful for natural tasks that require knowledge of the **salient attributes** of the data, which include tasks like face recognition and object recognition.

> **In this paper, rather than using a single unstructured noise vector, we propose to decompose the input noise vector into two parts: (1) z, which is treated as source of incompressible noise; (2) c, which we will call the latent code and will target the salient structured semantic features of the data distribution.**

> We provide the generator network with both the incompressible noise z and the latent code c, so the form of the generator becomes $G(z, c)$. **In standard GAN, the generator is free to ignore the additional latent code c by finding a solution satisfying $P_G(x|c) = P_G(x)$.**

> **We propose an information-theoretic regularization: there should be high mutual information between latent codes c and generator distribution $G(z, c)$. Thus $I(c; G(z, c))$ should be high.**

> In information theory, mutual information between X and Y, $I(X, Y)$, **measures the "amount of information" learned**

> **from knowledge of random variable Y about the other random variable X.**

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

> **I(X;Y) is the reduction of uncertainty in X wher Y is observed. If X and Y are independent, then I(X;Y)=0; by contrast, if X and Y are related by a determniistic, invertible function, then maximal mutual information is attained. This interpretation makes it easy to formulate a cost: given and $x \sim P_G(x)$, we want $P_G(c|x)$ to have a small entropy. In other words, the information in the latent code c should not be lost in the generation process.**

The loss function of this paper:

$$\min_G \max_D V_I(D,G) = V(D,G) - \lambda I(c; G(z,c))$$