Conditional Generative Adversarial Nets

Author: Zhang Xiaozheng Date: 2019/11/28

In generative adversarial nets, we can generate the pictures which is very realistic. But there is some problems which is we can't control the mode of these generative pictures.

So to get the picture we want, this paper use the conditional generative adversarial nets. In generative adversarial nets, we change the function to

$$min_G max_D \ V(G,D) = E_{x \sim p_{data}}[log(D(x|y))] + E_{z \sim p_z}[log(1 - log(G(z|y)))]$$

So this paper simply fed the label y to the nets. In the generator net, a noise prior z with dimensionality 100 was drawn from a uniform distribution within the unit hypercube. Both z and y are mapped to hidden layers with Rectified Linear Unit (ReLu) activation, with layer sizes 200 and 1000 respectively, before both being mapped to second, combined hidden ReLu layer of dimensionality 1200. Then the net have a final sigmoid unit layer as the output for generating the 784-dimensional MNIST samples.

Besides, this paper also propose a new way to tag pictures automatically. The idea is let the pictures to be condition and tags to be outputs. But the problem is how to represent words and pictures? The way this paper give is to use word2vec and a pre-trained convolution nets. So the output of this convolution nets is the pictures and generate word vectors.

For each image, they generate 100 samples and select top 20 closest words using cosine similarity. In this way, we can tag this picture with more than one tag.

- Multi-modal Learning

In real word, there are many kinds of message such as texts, pictures, sound...... We human being learn the new knowledge by accept these kind of message. So we hope the machine to learn more than one kind of message. An example is describe a given picture.

To we process the picture message, we usually use the CNN and for the text, we usually use RNN. This paper give a novel way to do multi-modal learning.