4712099

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention

1. The neural network captures the salient features of an image in the lower level representation of the CNN and uses either of the *deterministic attention* mechanism trained by back propagation methods or uses *stochastic attention* trained by maximizing an approximate variational lower bound. Which of the following is true about the attention mechanism specified?

   a. The Stochastic (Hard) attention mechanism is a REINFORCE learning rule where a objective function L_s is approximated by applying gradient where word is generated at every time step t dependent on hidden state and other sequence of words previously generated (probability that location I is the right place to focus to produce the next word).
   b. The Deterministic (Soft) attention mechanism gives attention to a location i in the image in blending the annotation vectors a_i's drawn from different locations together to produce the next word in forming the caption of the input image.
   c. Either a or b
   d. Only a seems feasible

   Answer: c

WAVENET: A GENERATIVE MODEL FOR RAW AUDIO

1. WaveNet, a deep neural network, is modeled based on PixelCNN architecture. This model *generates* new and highly realistic music/audio fragments. Which of the following is true about the WaveNet model?

   a. Each audio sample x_t is conditional probability distribution based on the previous samples of previous timestamps (stack of convolutional layers) and not on the future timestamps
   b. Pooling is the most important part in the CNN network used by the model and not dilated casual convolutions (convolution with holes) to increase the receptive field by orders of magnitude
   c. The output of the model had the same time dimensionality as the input and is a categorical distribution over the next value x_i with a softmax layer. Additionally, overfitting and under fitting can be tested by tuning the hyper parameters.
   d. Only a and c

   Answer: d