
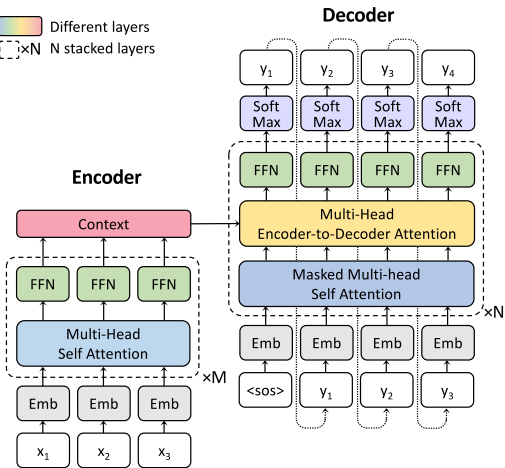
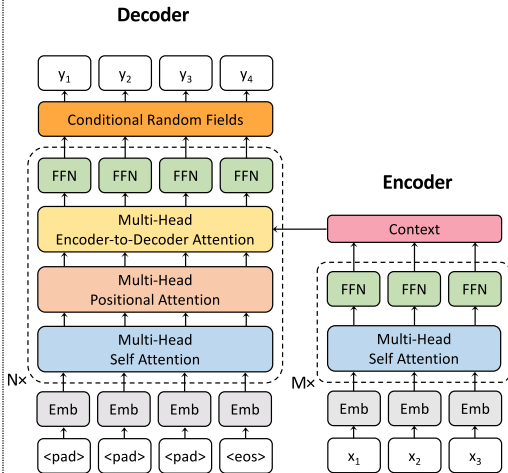
 Different layers  
  $\times N$  N stacked layers



(a) Autoregressive Transformer



(b) Non-Autoregressive Transformer with Conditional Random Fields

Figure 1: Illustration of the Transformer model and our Transformer-based NART-CRF model