Attention Mechanism in Equations (2)

Decoder state:

$$s_i = \tanh(U_d s_{i-1} + W_d E_d \hat{y}_{i-1} + C c_i + b_d)$$

Output projection:

$$t_i = \tanh\left(U_o s_i + W_o E_d \hat{y}_{i-1} + C_o c_i + b_o\right)$$

...context vector is mixed with the hidden state

Output distribution:

$$p\left(y_i = k \mid s_i, y_{i-1}, c_i\right) \propto \exp\left(W_o t_i\right)_k + b_k$$