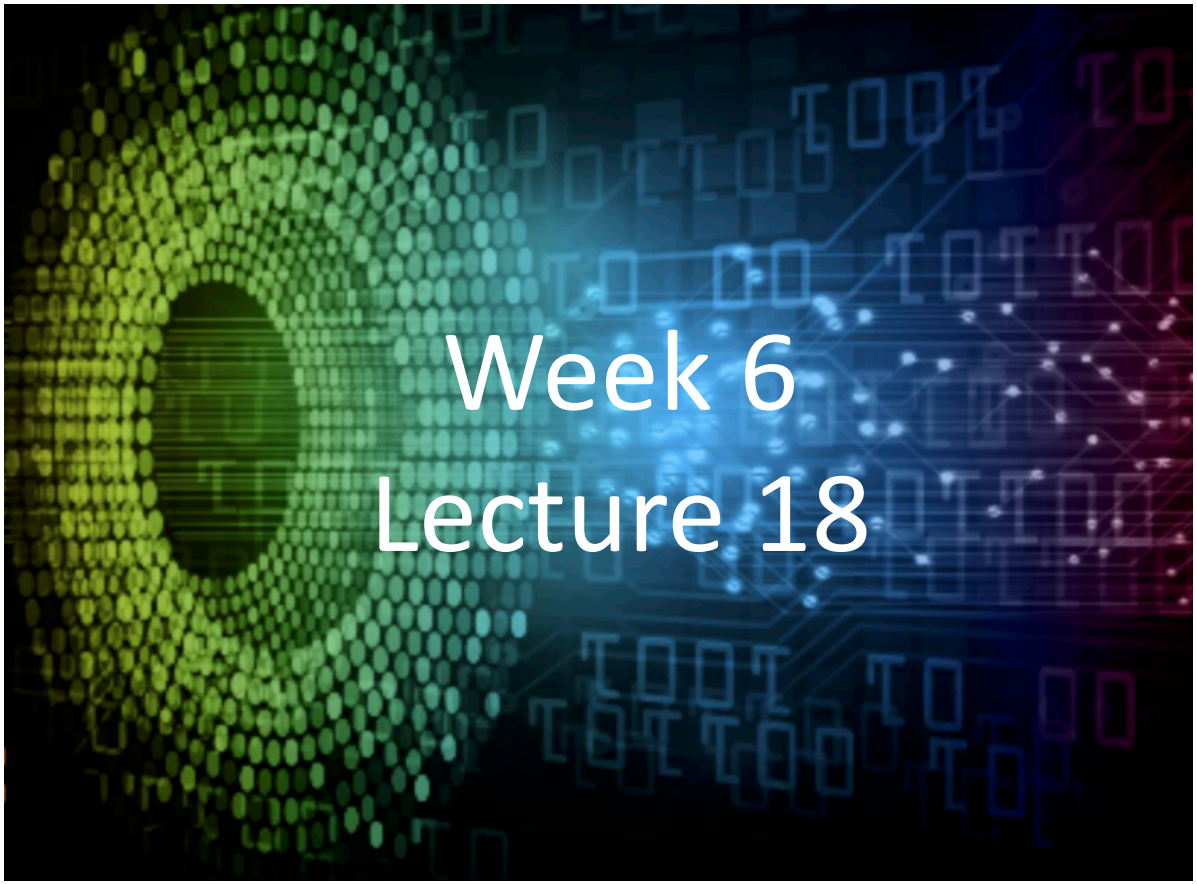
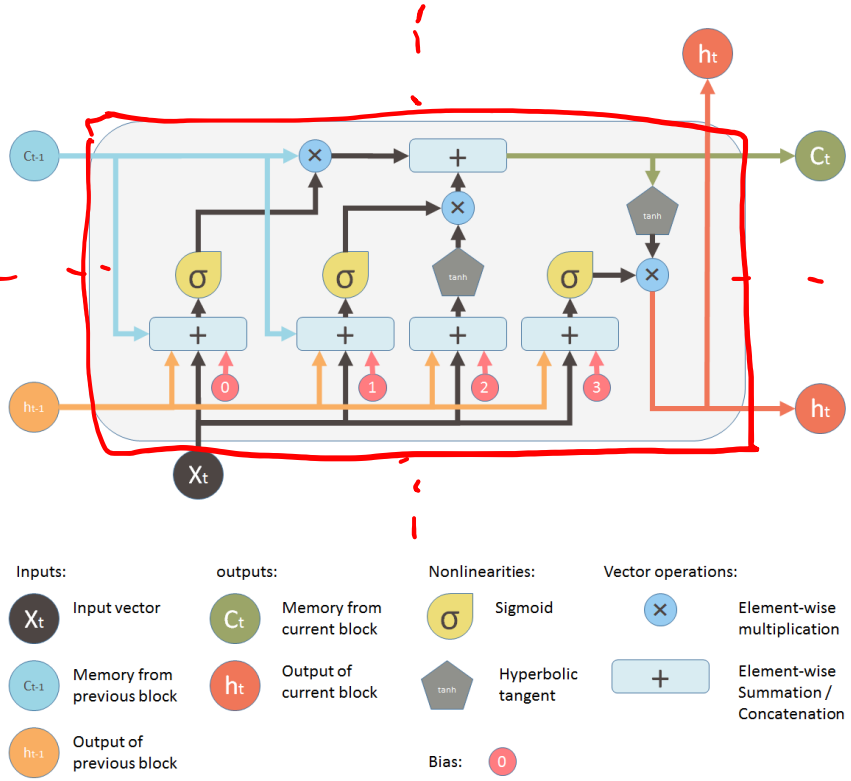
Introduction to Deep Learning
Applications and Theory



ECE 596 / AMATH 563

Previous Lecture: Recurrent Neural Networks (RNNs)

- Diminishing/ Exploding Gradients
- Gated RNNs
  - GRU
  - LSTM



This Lecture: Architectures of RNN Blocks

- Bi-Directional RNN
- Deep RNN
- Encoder / Decoder RNN
- Sequence to Sequence
- Attention

LSTM

$$\tilde{c}^{<t>} = \tanh(W_c[a^{<t-1>}; x^{<t>}] + b_c)$$

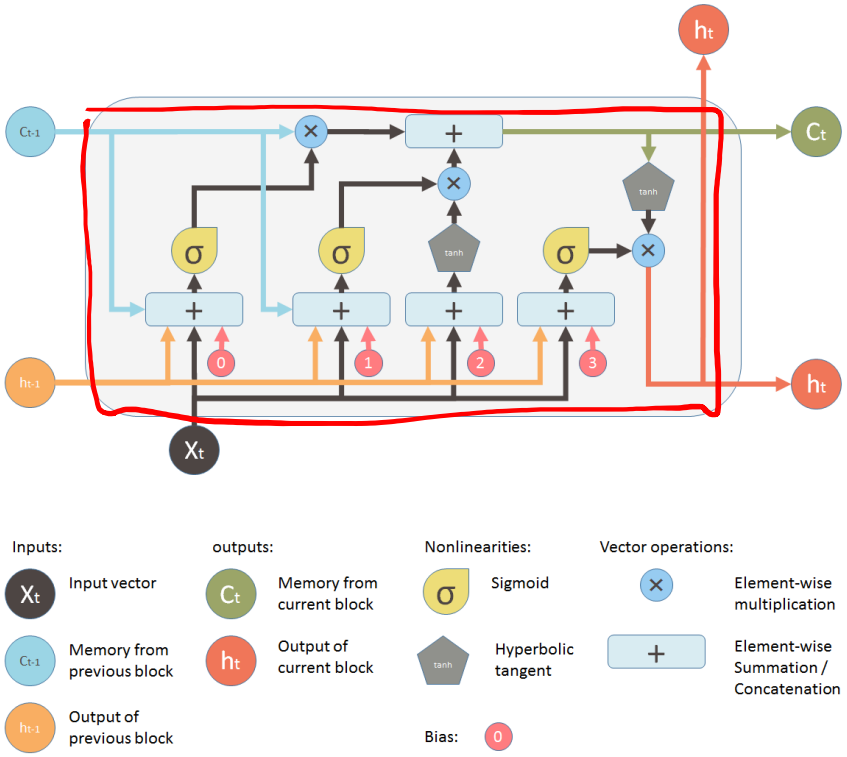
$$\Gamma_u = \sigma(W_u[a^{<t-1>}; x^{<t>}] + b_u)$$

$$\Gamma_f = \sigma(W_f[a^{<t-1>}; x^{<t>}] + b_f)$$

$$\Gamma_o = \sigma(W_o[a^{<t-1>}; x^{<t>}] + b_0)$$

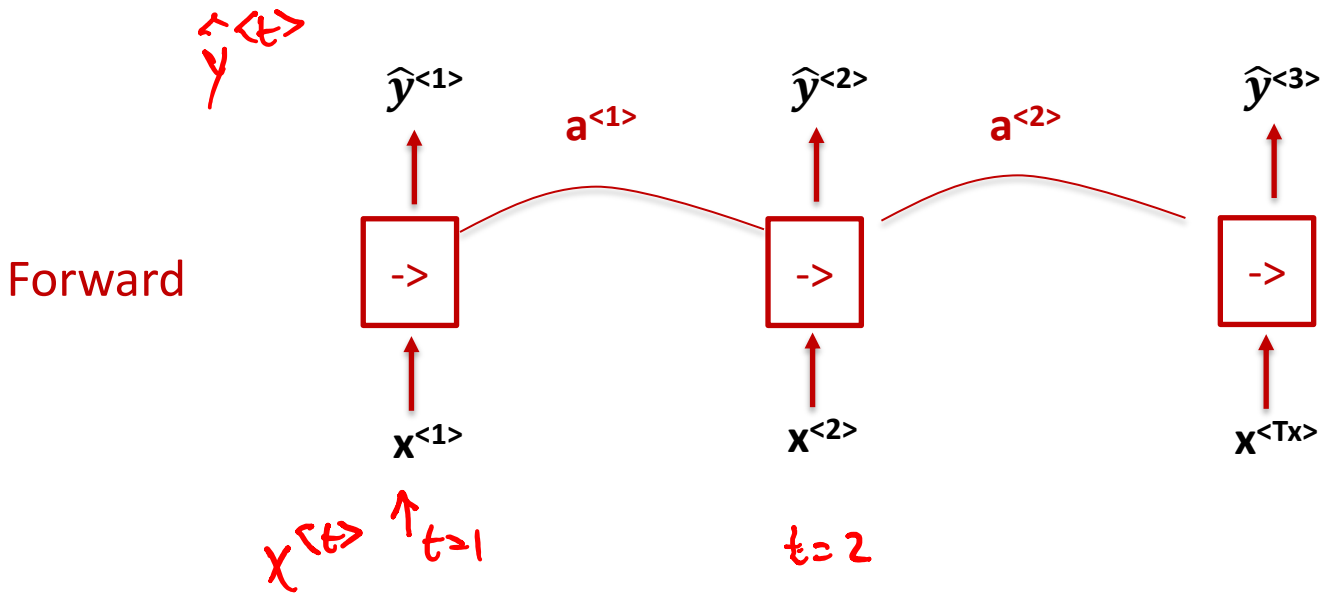
$$c^{<t>} = \Gamma_u \tilde{c}^{<t>} + \Gamma_f c^{<t-1>}$$

$$a^{<t>} = \Gamma_o \tanh c^{<t>}$$



Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

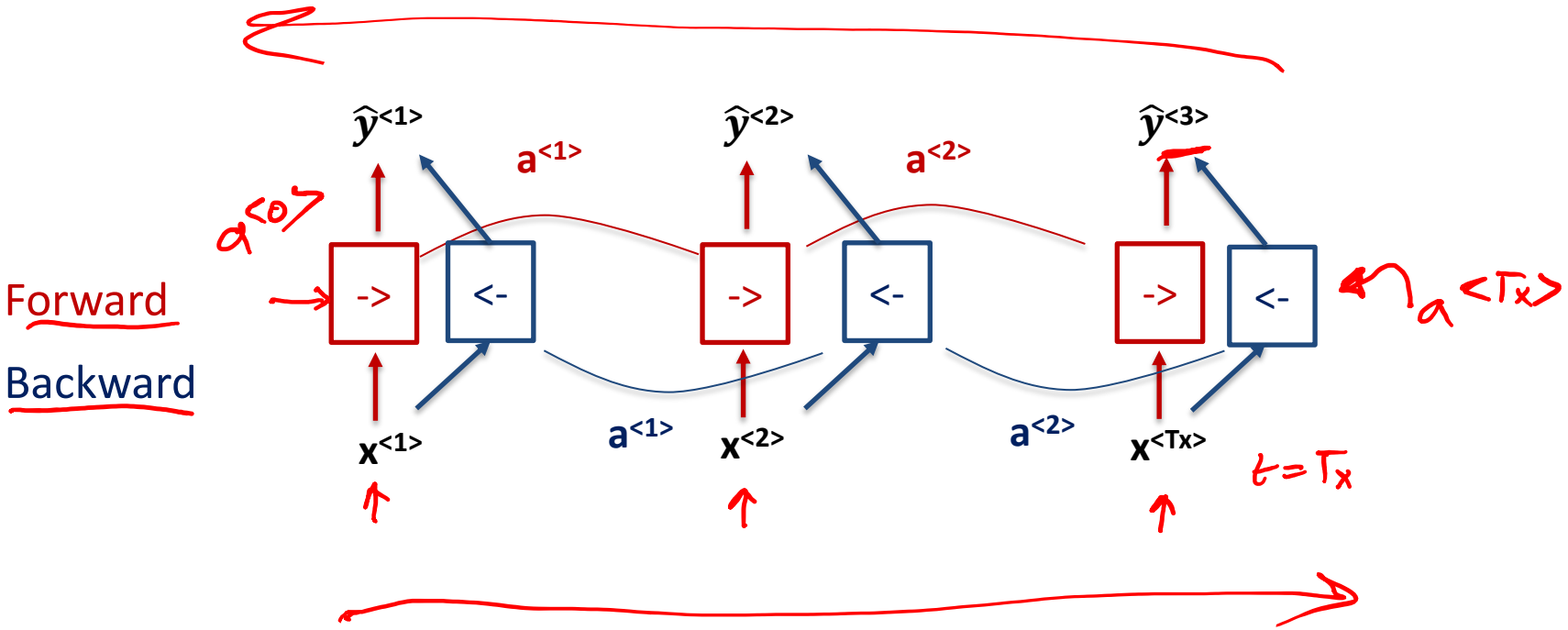
Bi-Directional RNN



He said, Tesla is a unit of magnetic field strength

He said, Tesla is an electric automotive and sustainable energy company

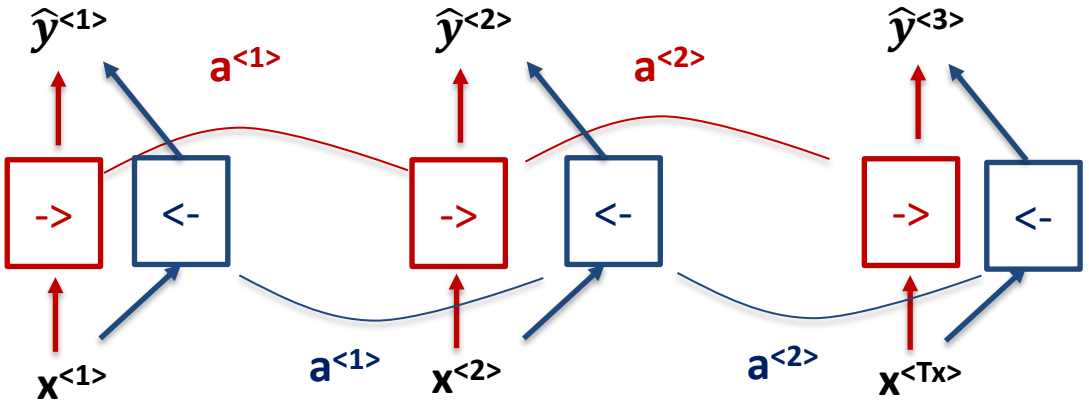
Bi-Directional RNN



Bi-Directional RNN

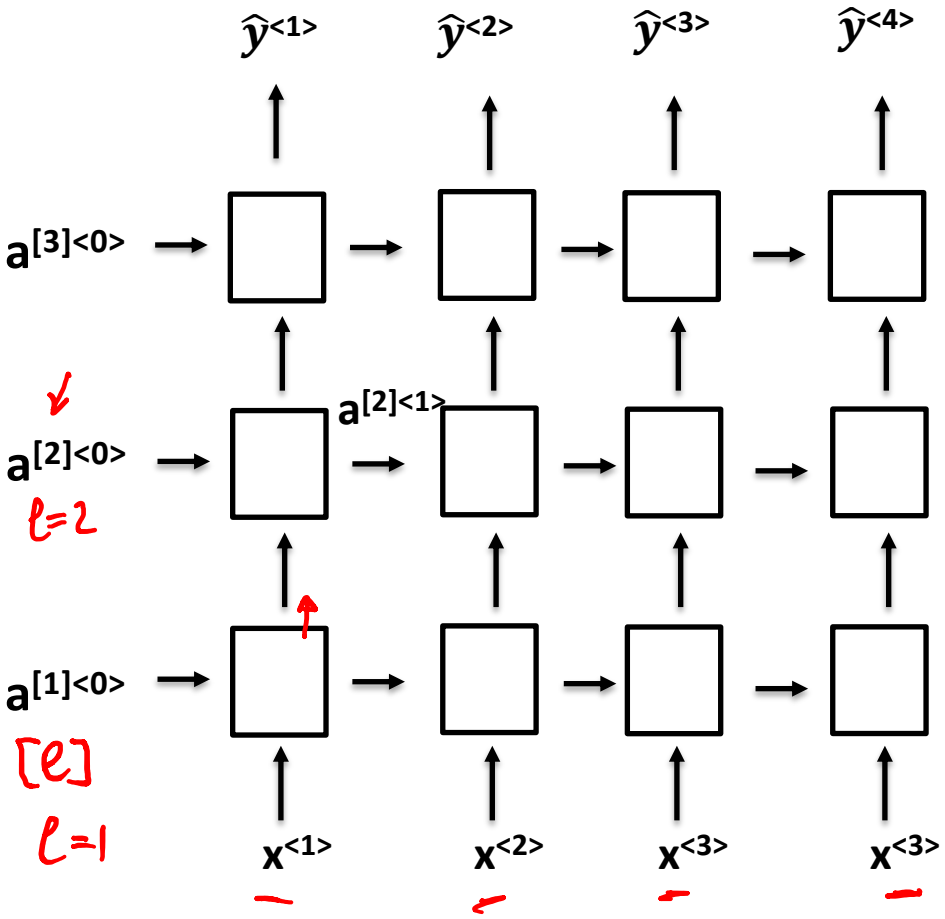
Forward

Backward

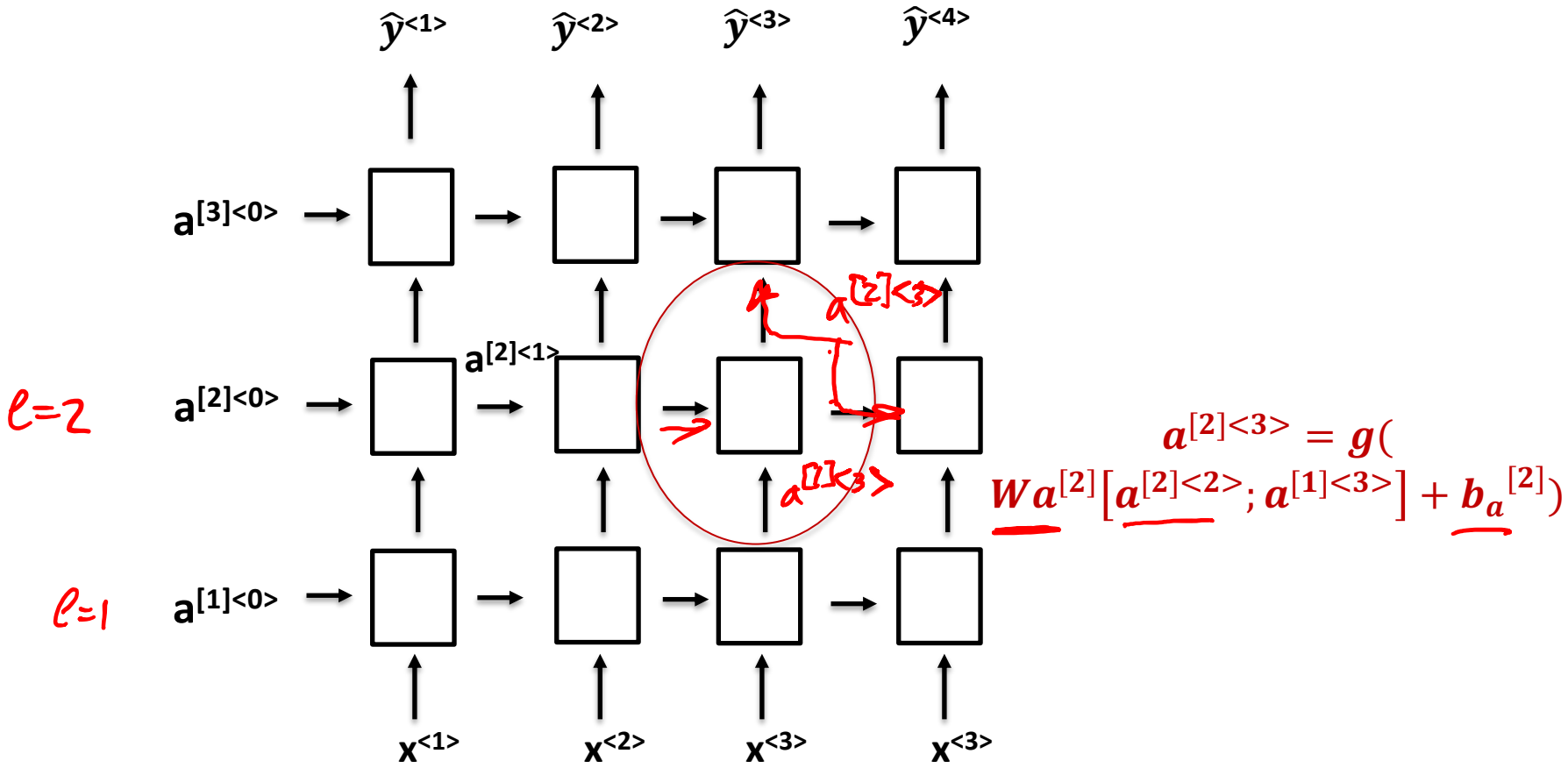


$$\hat{y}^{<t>} = g(Wy\,[a^{F<t>}; a^{B<t>}] + by)$$

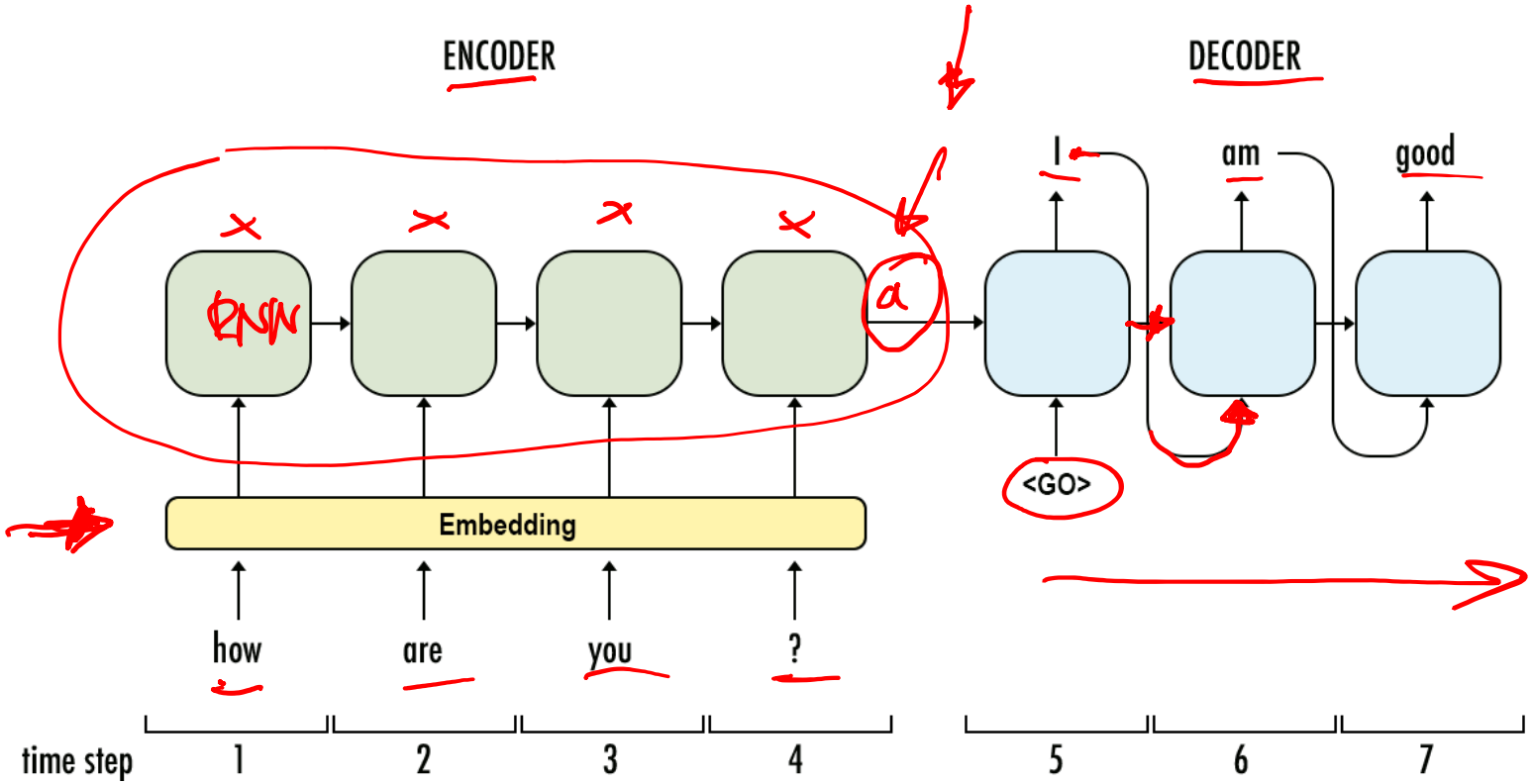
Deep RNN



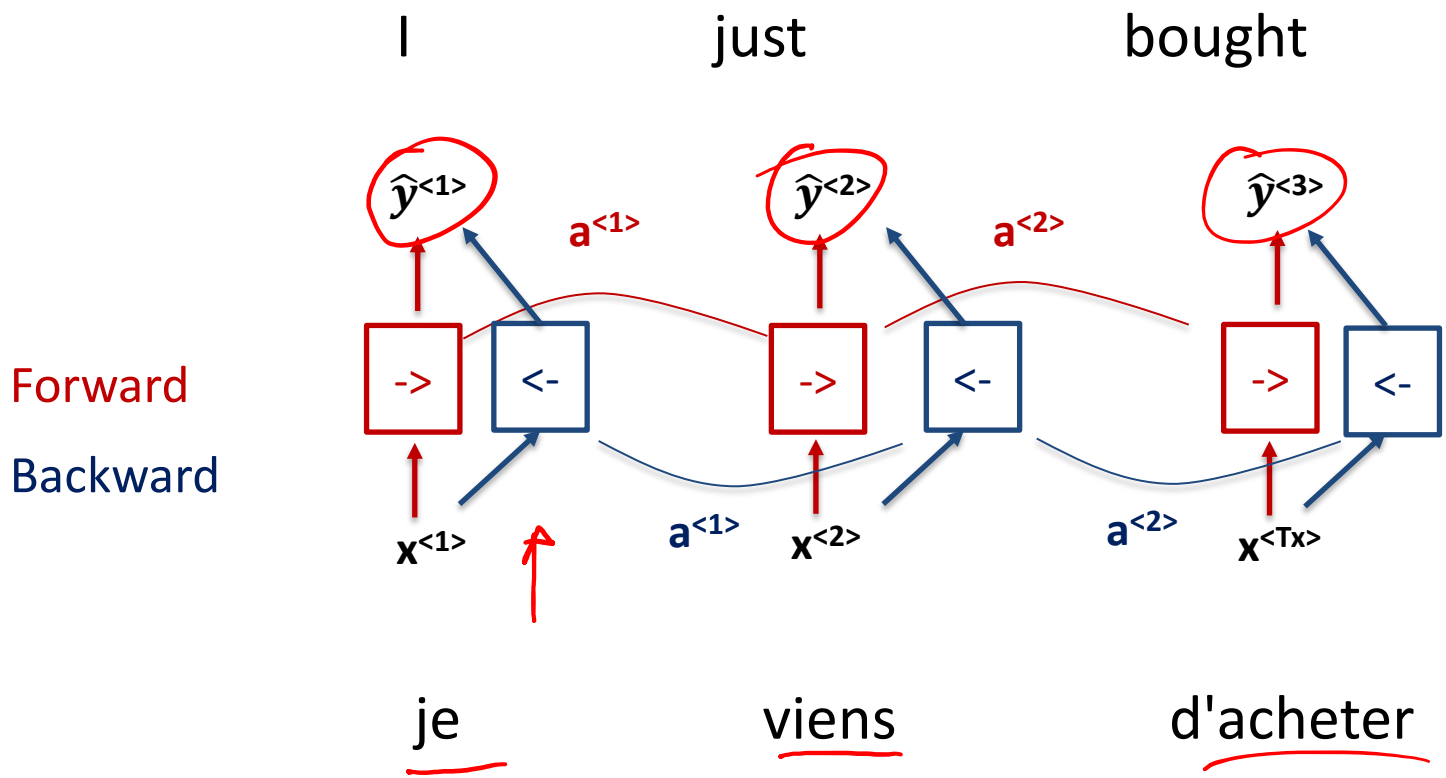
Deep RNN


$\hat{y}^{<1>}$
$\hat{y}^{<2>}$
$\hat{y}^{<3>}$
$\hat{y}^{<4>}$
$a^{[3]<0>}$
$a^{[2]<0>}$
$a^{[2]<1>}$
$a^{[1]<0>}$
$x^{<1>}$
$x^{<2>}$
$x^{<3>}$
$x^{<3>}$
$a^{[2]<3>}$
$a^{[1]<3>}$
$\ell=2$
$\ell=1$
$a^{[2]<3>} = g(W a^{[2]}[a^{[2]<2>}; a^{[1]<3>}] + b_a{}^{[2]})$

Seq2Seq – Encoder Decoder

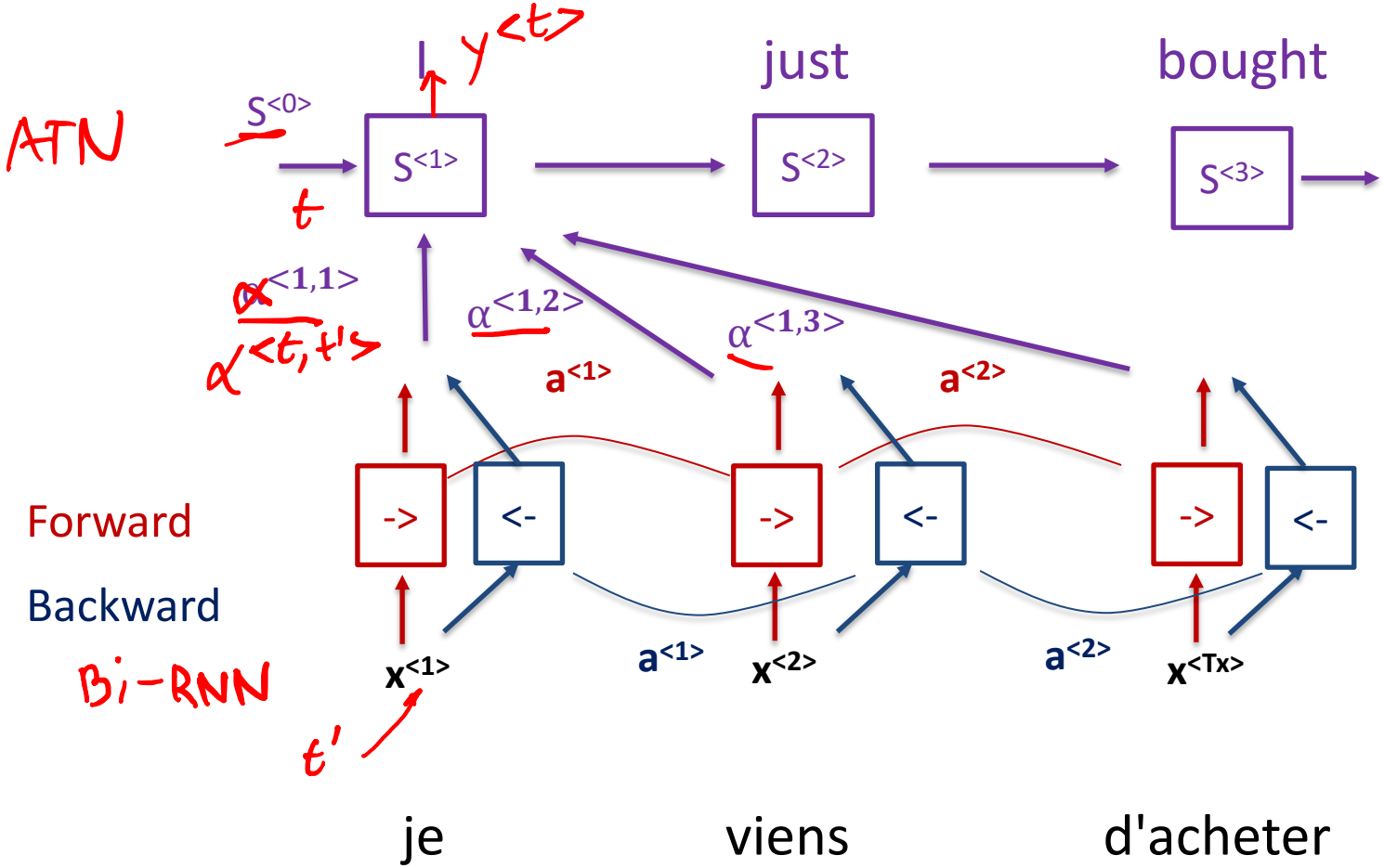


Attention Model



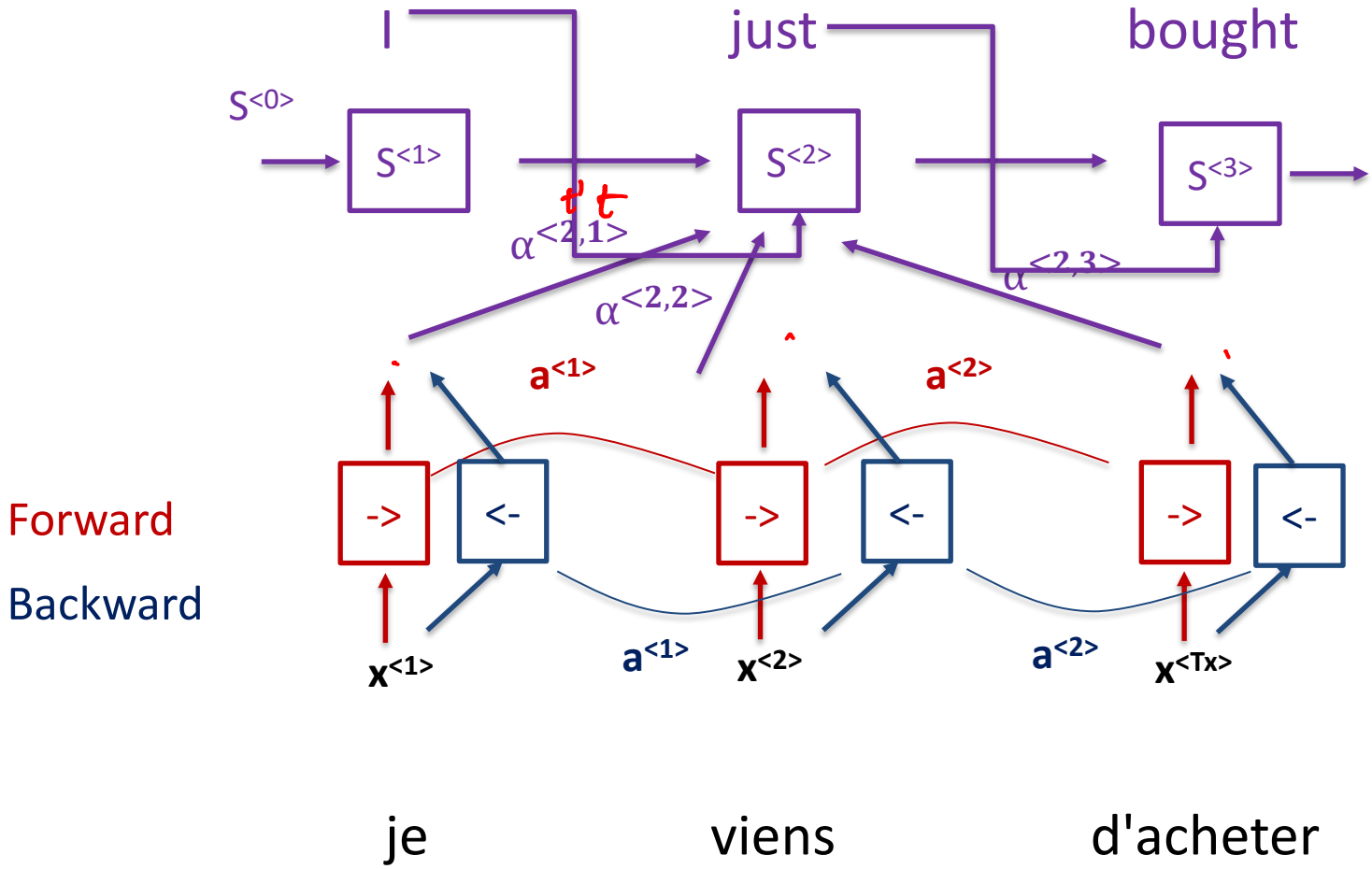
Bahdanau et al. "Neural machine translation by jointly learning to align and translate." (2014).

Attention Model



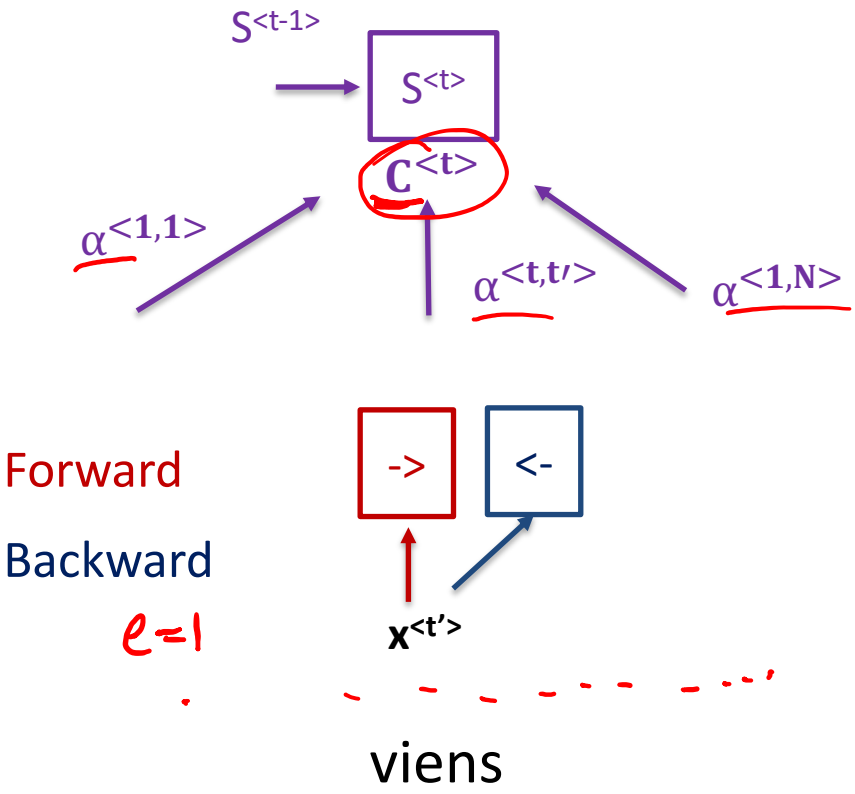
Bahdanau et al. "Neural machine translation by jointly learning to align and translate." (2014).

Attention Model



Bahdanau et al. "Neural machine translation by jointly learning to align and translate." (2014).

Attention Model



$$C^{<t>} = \sum_{t'} \alpha^{<t,t'>} a^{<t'>}$$

FC

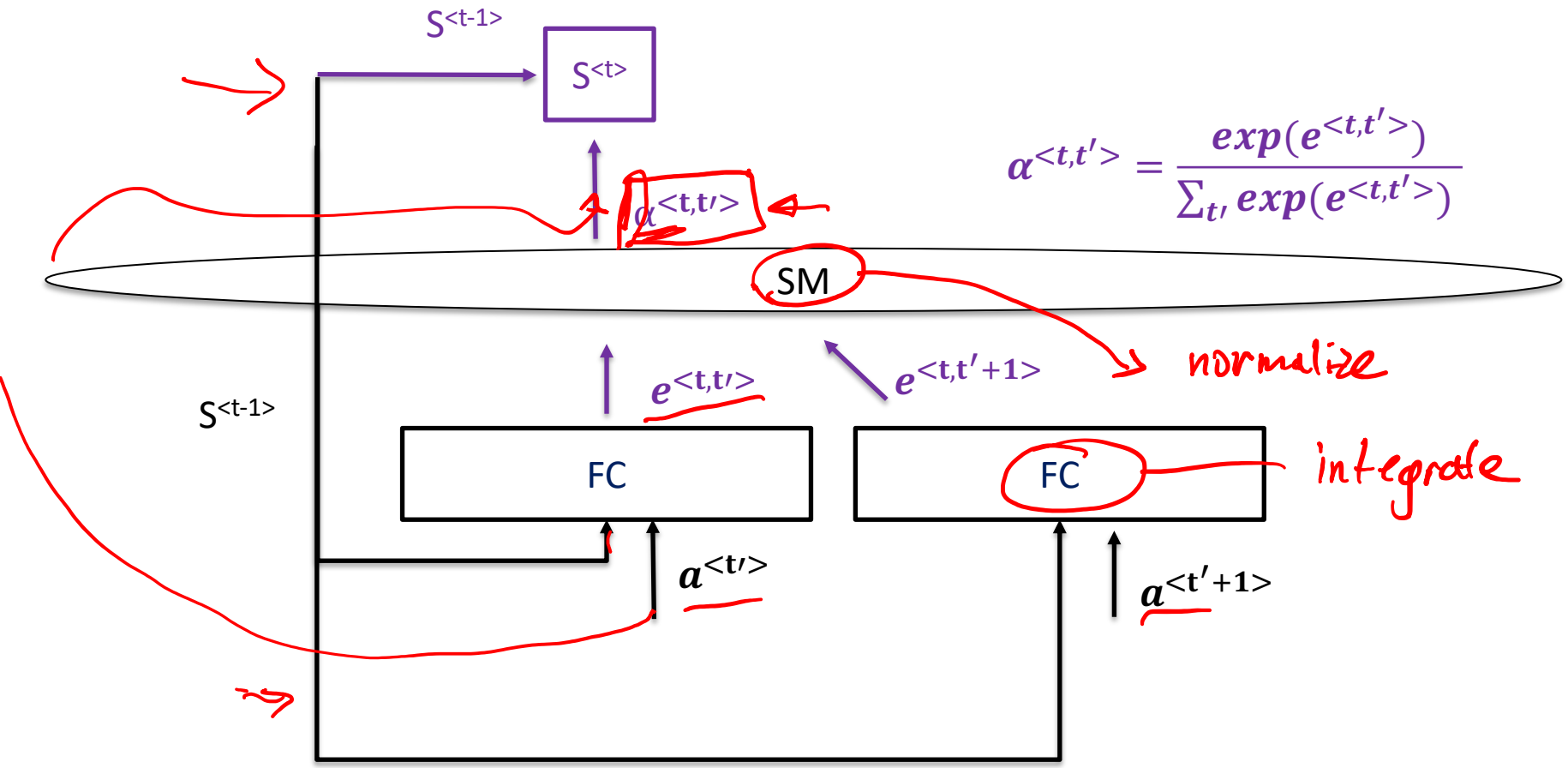
$$\sum_{t'} \alpha^{<t,t'>} = 1$$

SM

$$a^{<t'>} = (a^{<t'>}, a^{<t'>})$$

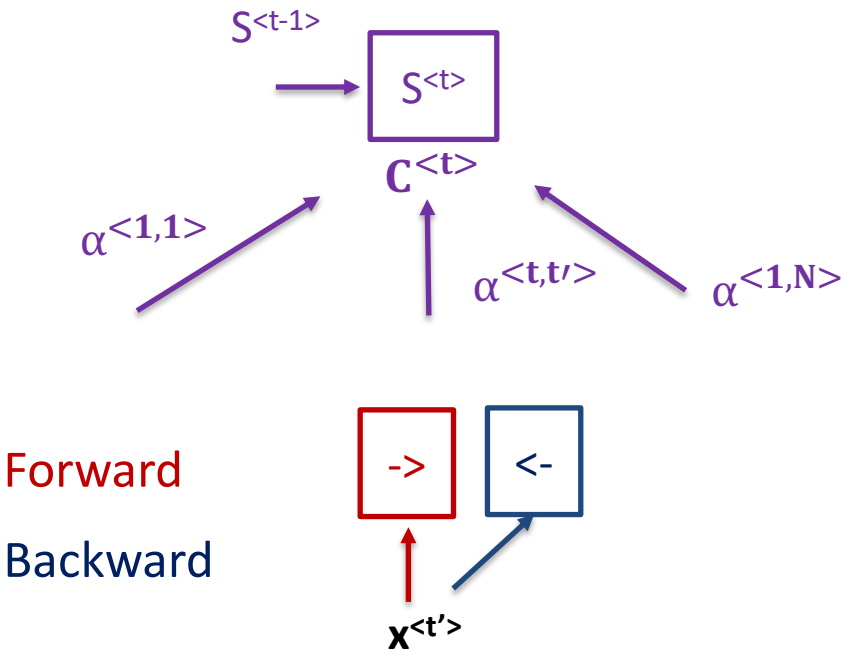
Bahdanau et al. "Neural machine translation by jointly learning to align and translate." (2014).

Attention Model



$$\alpha^{<t,t'>} = \frac{exp(e^{<t,t'>})}{\sum_{t'} exp(e^{<t,t'>})}$$

Attention Model



$$C^{<t>} = \sum_{t'} \alpha^{<t,t'>} a^{<t'>}$$

$$\sum_{t'} \alpha^{<t,t'>} = 1$$

$$a^{<t'>} = (a^{<t'>}, a^{<t'>})$$

Bahdanau et al. "Neural machine translation by jointly learning to align and translate." (2014).