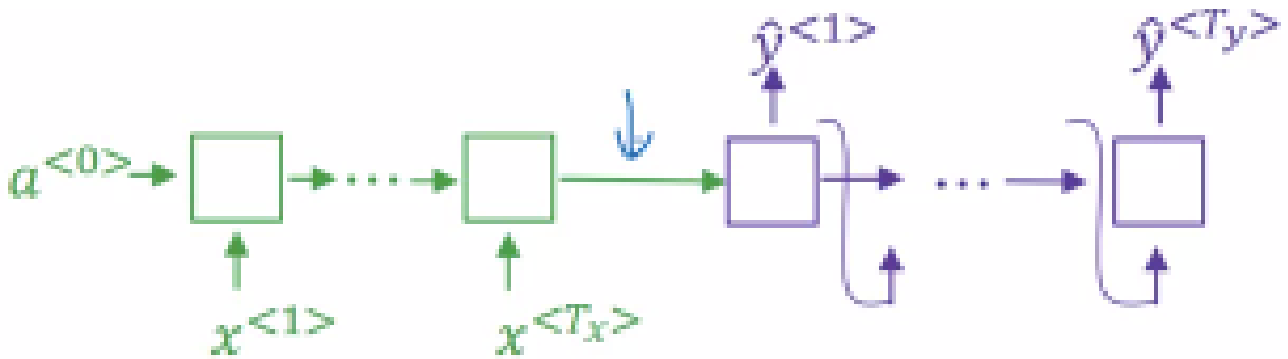
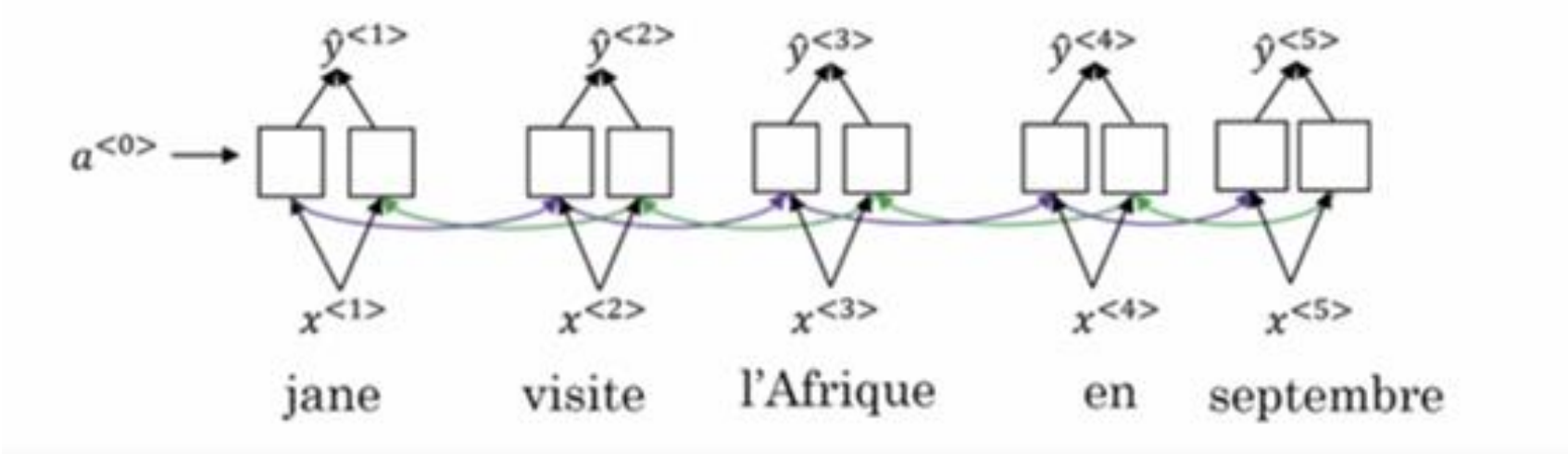
Attention Model

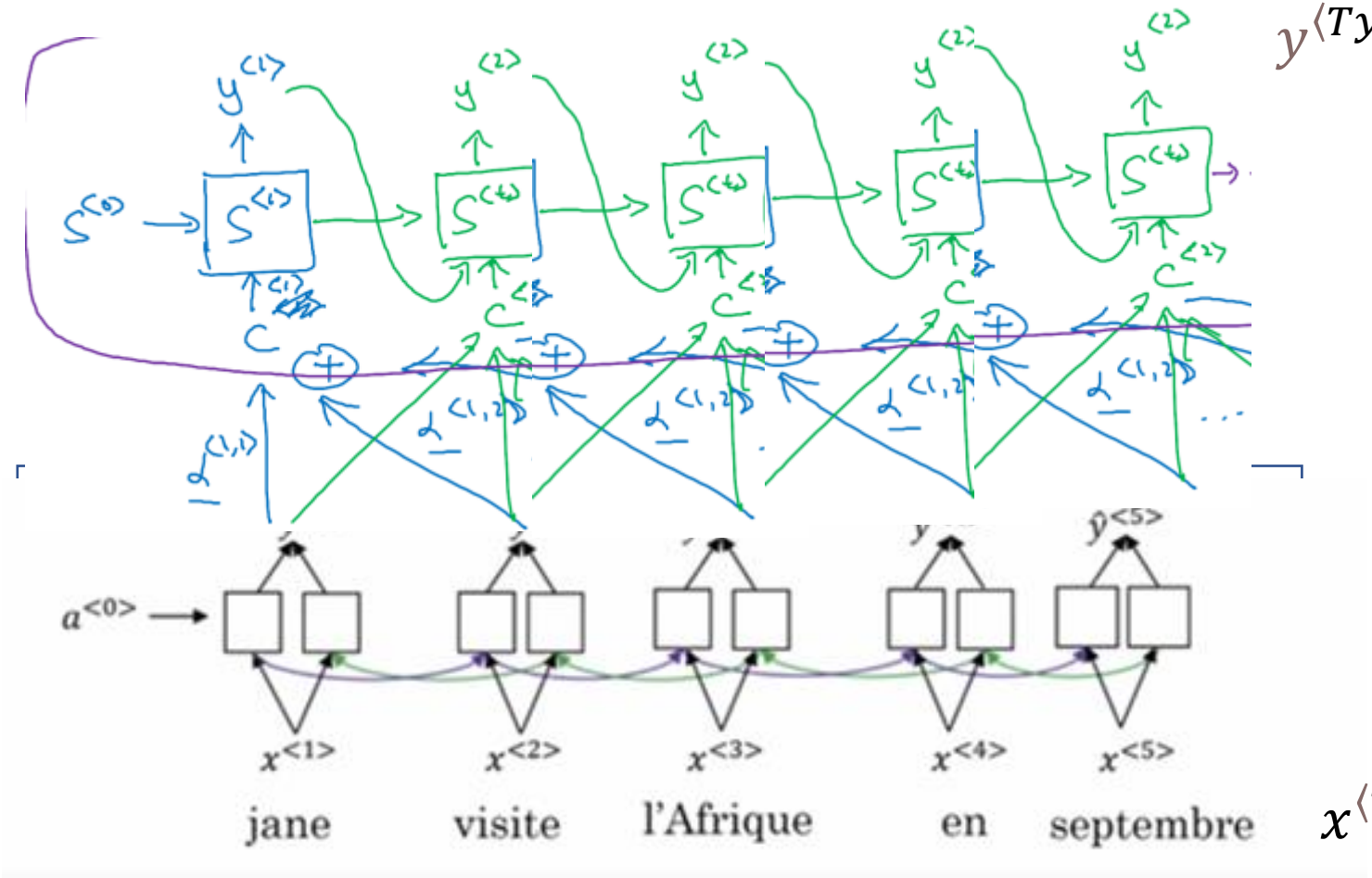
26 June 2021 Yoon Choi

Jane went to Africa last September, and enjoyed the culture and met many wonderful people; she came back raving about how wonderful her trip was, and is tempting me to go too.



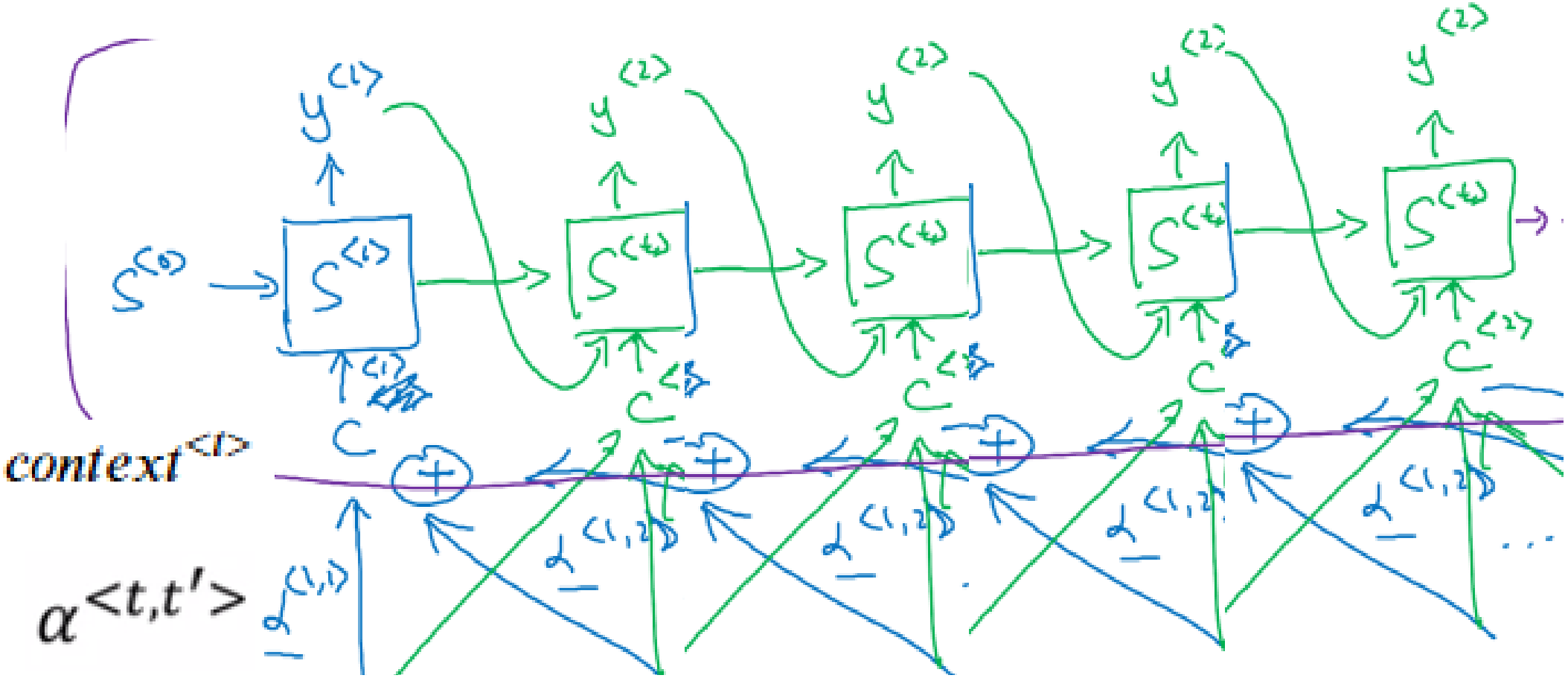
Jane s'est rendue en Afrique en septembre dernier, a apprécié la culture et a rencontré beaucoup de gens merveilleux; elle est revenue en parlant comment son voyage était merveilleux, et elle me tente d'y aller aussi.

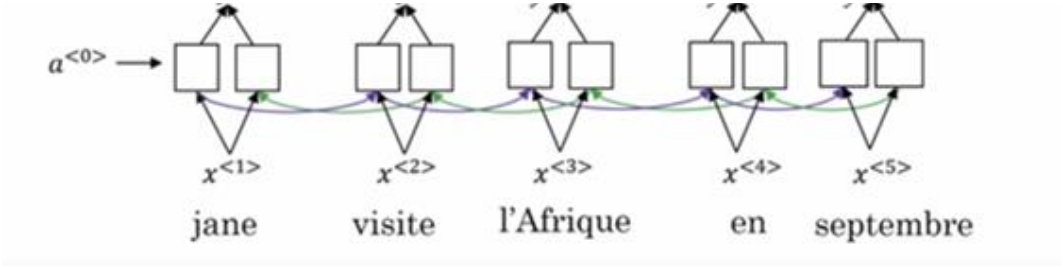

$\hat{y}^{<1>}$
$\hat{y}^{<2>}$
$\hat{y}^{<3>}$
$\hat{y}^{<4>}$
$\hat{y}^{<5>}$
$a^{<0>}$
$x^{<1>}$
$x^{<2>}$
$x^{<3>}$
$x^{<4>}$
$x^{<5>}$
jane
visite
l'Afrique
en
septembre



$y^{\langle Ty \rangle}$ < −output

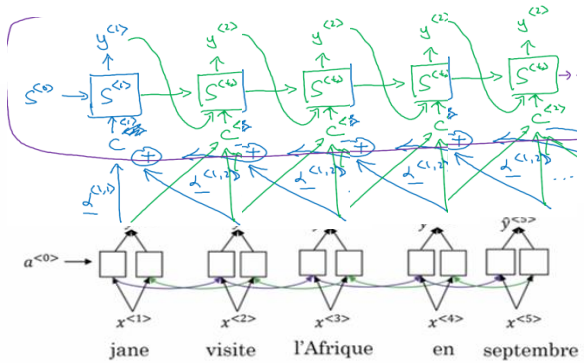
$x^{\langle Tx \rangle}$ <- input


context<1>
α<t,t'>

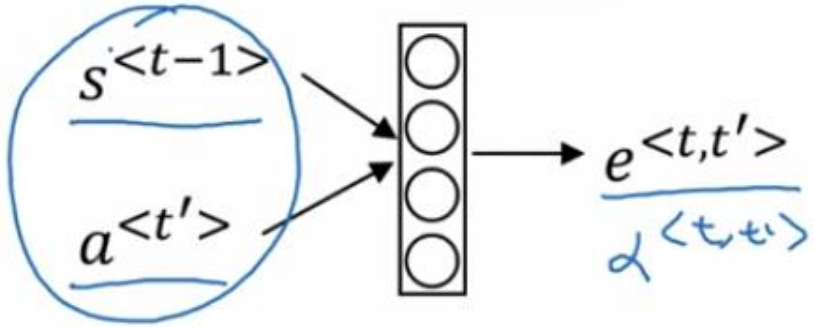

a<0>
x<1>
x<2>
x<3>
x<4>
x<5>
jane
visite
l'Afrique
en
septembre

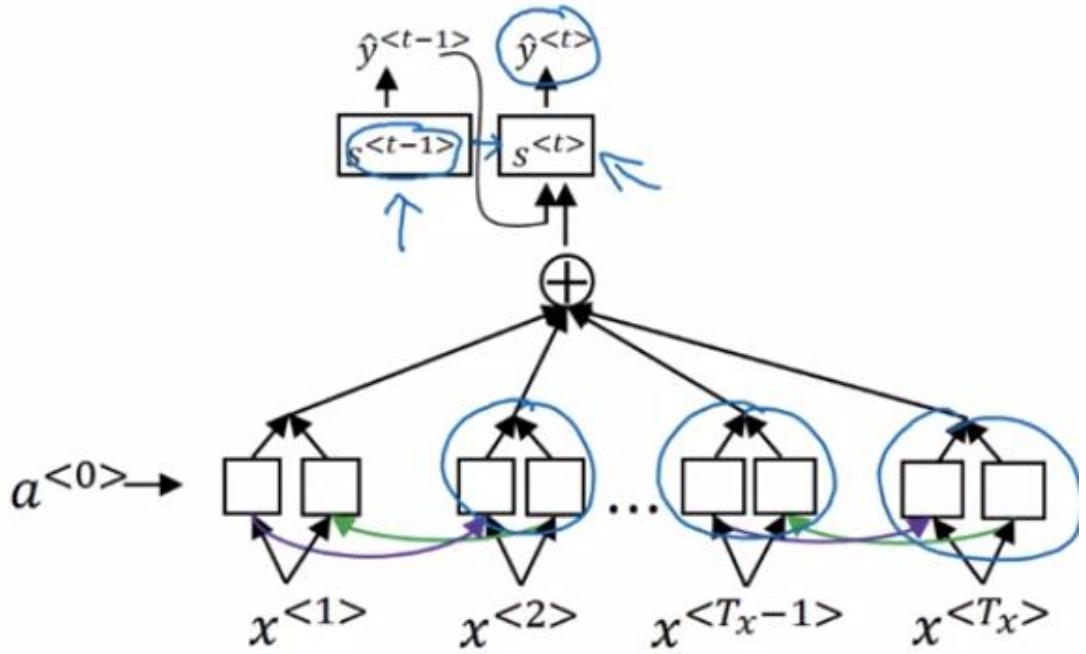
Computing attention $\alpha^{<t,t'>}$

$\alpha^{<t,t'>}$ = amount of attention $y^{<t>}$ should pay to $a^{<t'>}$

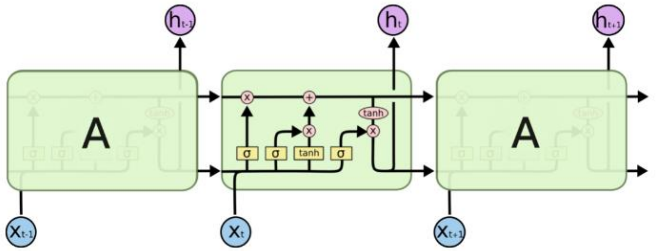


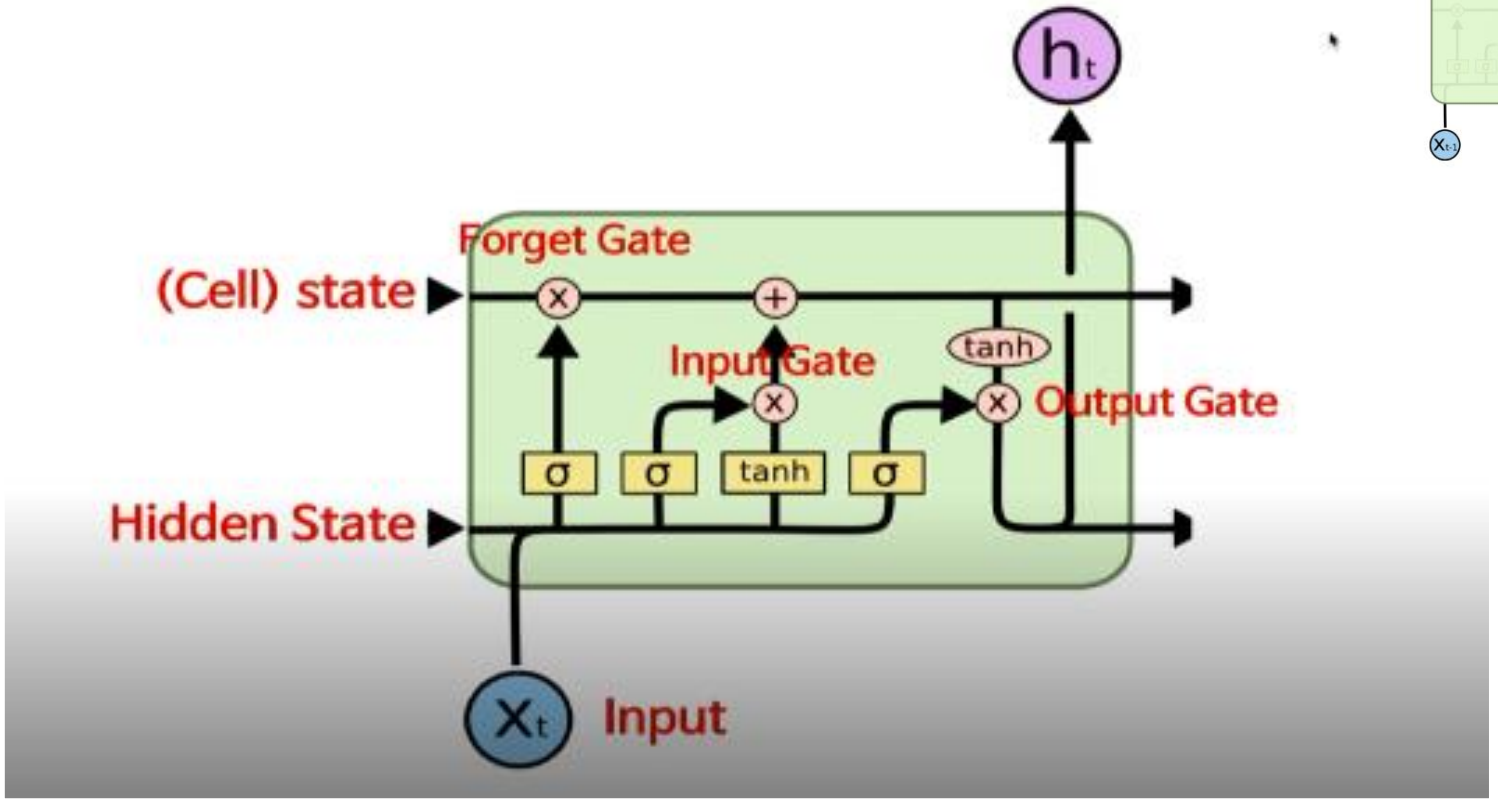
$$\alpha^{<t,t'>} = \frac{\exp(e^{<t,t'>})}{\sum_{t'=1}^{T_x} \exp(e^{<t,t'>})}$$

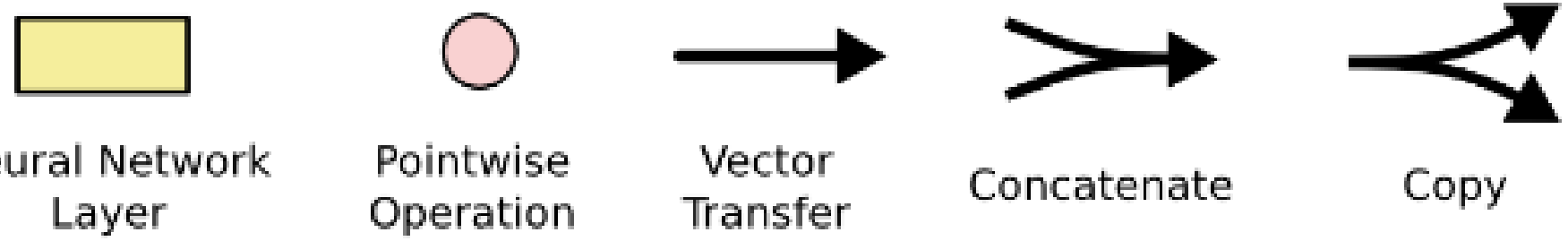


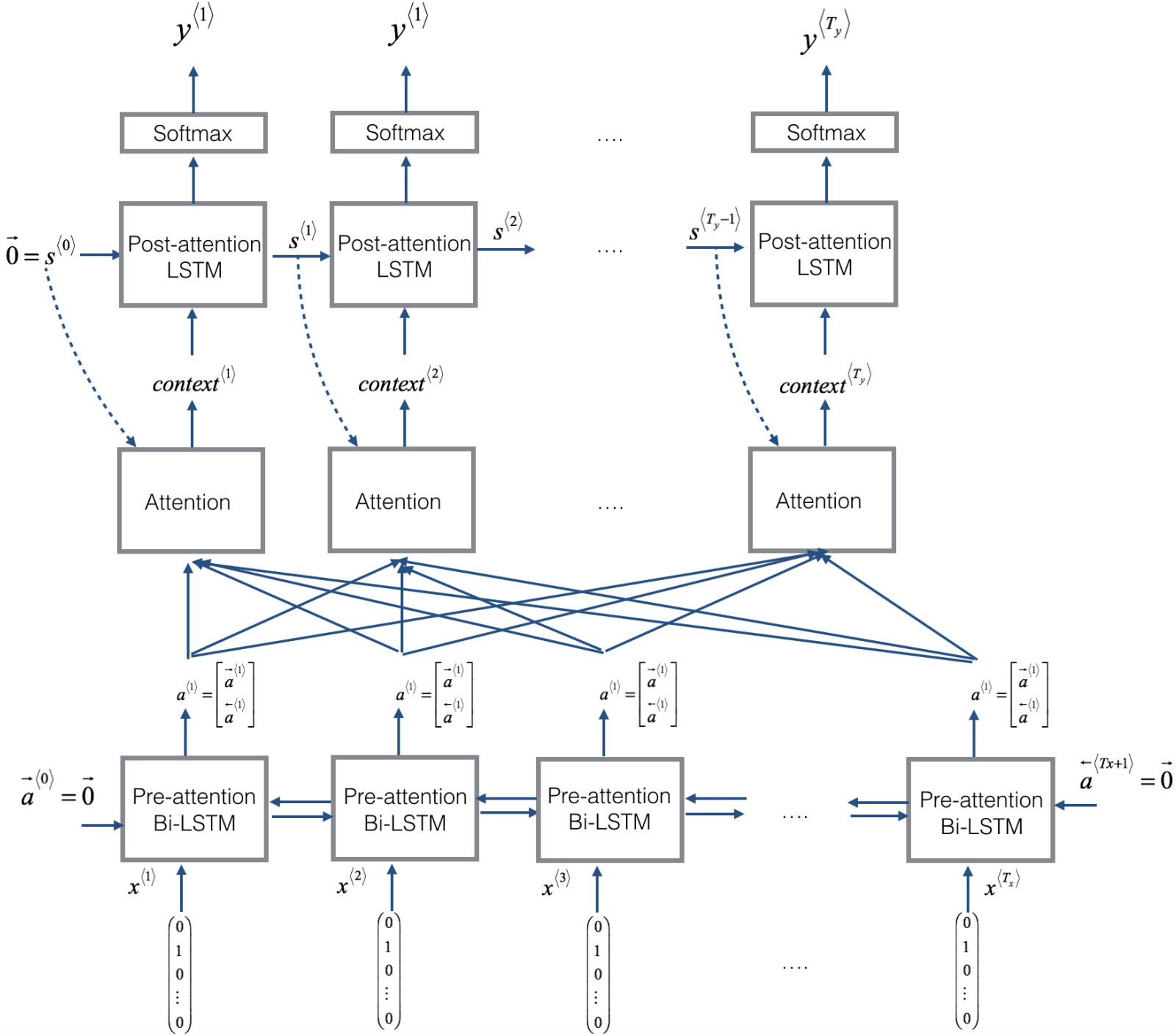


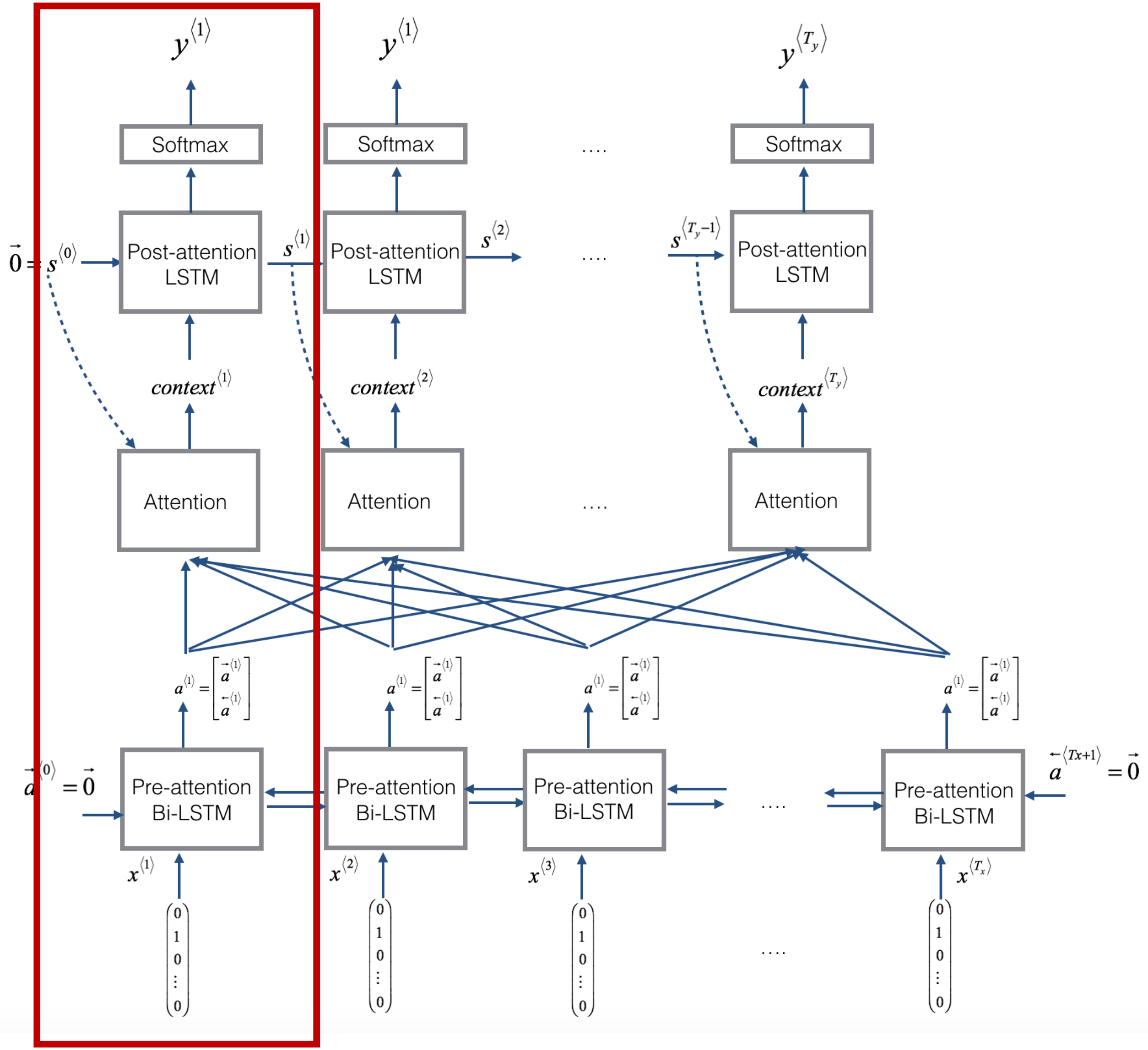
$$context^{<t>} = \sum_{t'=1}^{T_x} \alpha^{<t,t'>} a^{<t'>}$$


A
A

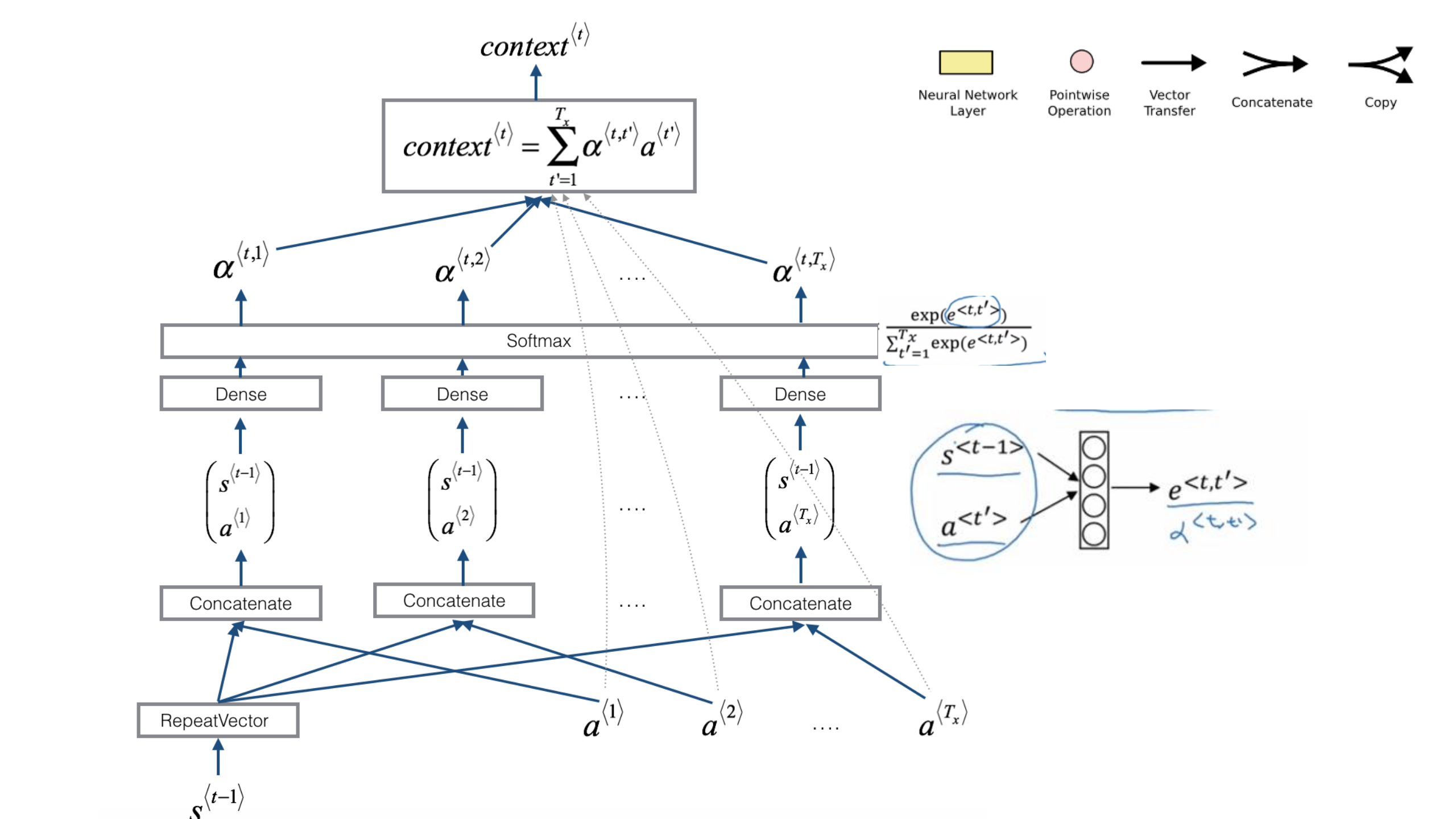

h_t
Forget Gate
(Cell) state
Input Gate
tanh
Output Gate
σ
σ
tanh
σ
Hidden State
X_t
Input


Neural Network Layer
Pointwise Operation
Vector Transfer
Concatenate
Copy


$y^{\langle 1 \rangle}$
$y^{\langle 1 \rangle}$
$y^{\langle T_y \rangle}$
Softmax
Softmax
Softmax
$\vec{0} = s^{\langle 0 \rangle}$
Post-attention LSTM
$s^{\langle 1 \rangle}$
Post-attention LSTM
$s^{\langle 2 \rangle}$
$s^{\langle T_y - 1 \rangle}$
Post-attention LSTM
$context^{\langle 1 \rangle}$
$context^{\langle 2 \rangle}$
$context^{\langle T_y \rangle}$
Attention
Attention
Attention
$a^{\langle 1 \rangle} = \begin{bmatrix} \vec{a}^{\langle 1 \rangle} \\ \overleftarrow{a}^{\langle 1 \rangle} \end{bmatrix}$
$\vec{a}^{\langle 0 \rangle} = \vec{0}$
Pre-attention Bi-LSTM
Pre-attention Bi-LSTM
Pre-attention Bi-LSTM
Pre-attention Bi-LSTM
$\overleftarrow{a}^{\langle Tx+1 \rangle} = \vec{0}$
$x^{\langle 1 \rangle}$
$x^{\langle 2 \rangle}$
$x^{\langle 3 \rangle}$
$x^{\langle T_x \rangle}$



Softmax
Softmax
Softmax
Post-attention LSTM
Post-attention LSTM
Post-attention LSTM
context
context
context
Attention
Attention
Attention
Pre-attention Bi-LSTM
Pre-attention Bi-LSTM
Pre-attention Bi-LSTM
Pre-attention Bi-LSTM


$context^{\langle t \rangle}$
$context^{\langle t \rangle} = \sum_{t'=1}^{T_x} \alpha^{\langle t,t' \rangle} a^{\langle t' \rangle}$
$\alpha^{\langle t,1 \rangle}$
$\alpha^{\langle t,2 \rangle}$
$\alpha^{\langle t,T_x \rangle}$
Softmax
$\frac{\exp(e^{<t,t'>})}{\sum_{t'=1}^{T_x} \exp(e^{<t,t'>})}$
Dense
Dense
Dense
$\begin{pmatrix} s^{\langle t-1 \rangle} \\ a^{\langle 1 \rangle} \end{pmatrix}$
$\begin{pmatrix} s^{\langle t-1 \rangle} \\ a^{\langle 2 \rangle} \end{pmatrix}$
$\begin{pmatrix} s^{\langle t-1 \rangle} \\ a^{\langle T_x \rangle} \end{pmatrix}$
Concatenate
Concatenate
Concatenate
RepeatVector
$s^{\langle t-1 \rangle}$
$a^{\langle 1 \rangle}$
$a^{\langle 2 \rangle}$
$a^{\langle T_x \rangle}$
$s^{<t-1>}$
$a^{<t'>}$
$e^{<t,t'>}$
$\alpha^{<t,t'>}$
Neural Network Layer
Pointwise Operation
Vector Transfer
Concatenate
Copy

output	AA	BB	CC	DD
AA	1	0	0	0
BB	0	1	0	0
CC	0	0	1	0
DD	0	0	0	1