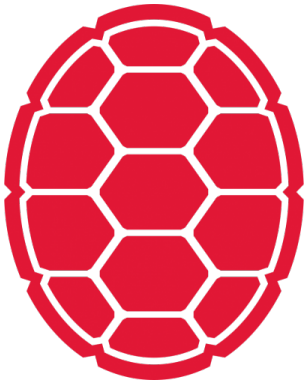
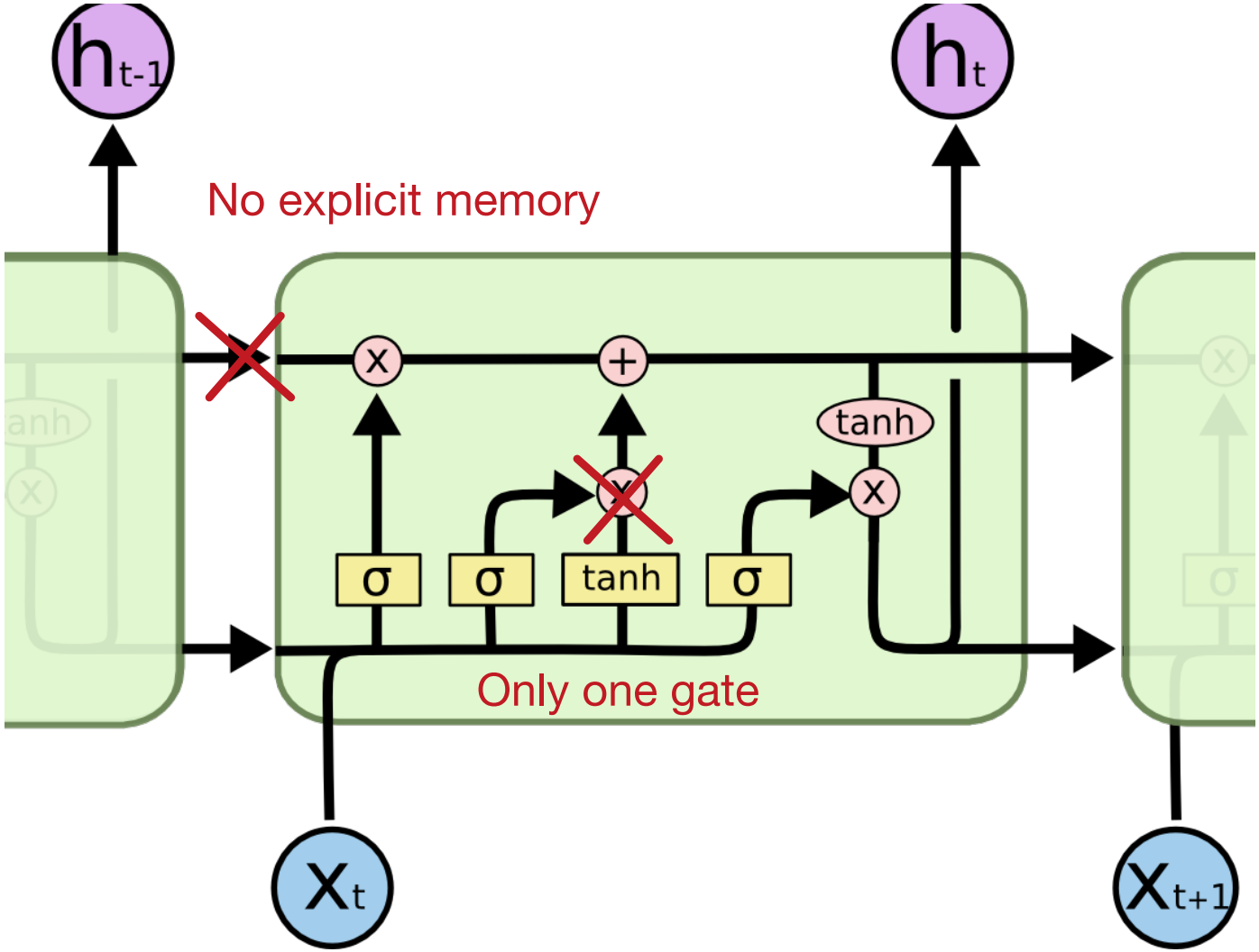


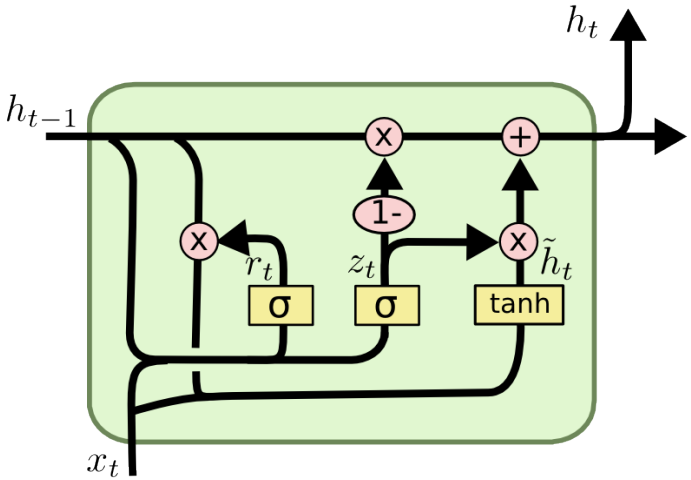
Sequence Models

Machine Learning: Jordan Boyd-Graber
University of Maryland
LSTM VARIANTS

GRU simplifies slightly



GRU simplifies slightly



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Slightly fewer parameters

What's most important part of LSTM

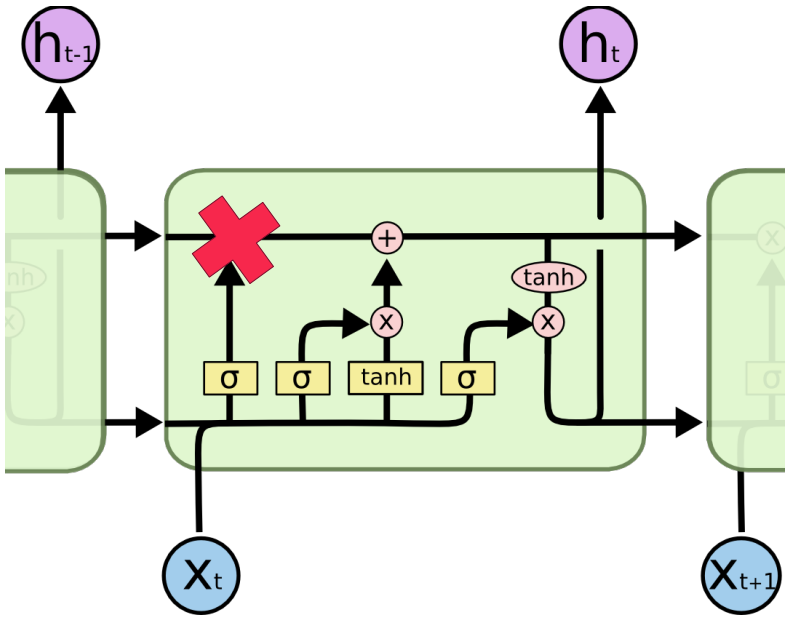
Greff et al. explore

- No Input Gate (NIG)
- No Forget Gate (NFG)
- No Output Gate (NOG)
- No Input Activation Function (NIAF)
- No Output Activation Function (NOAF)
- No Peepholes (NP)
- Coupled Input and Forget Gate (CIFG) : GRU, $f_t = 1 - i_t$
- Full Gate Recurrence (FGR): Original LSTM paper

What's most important part of LSTM

Greff et al. explore

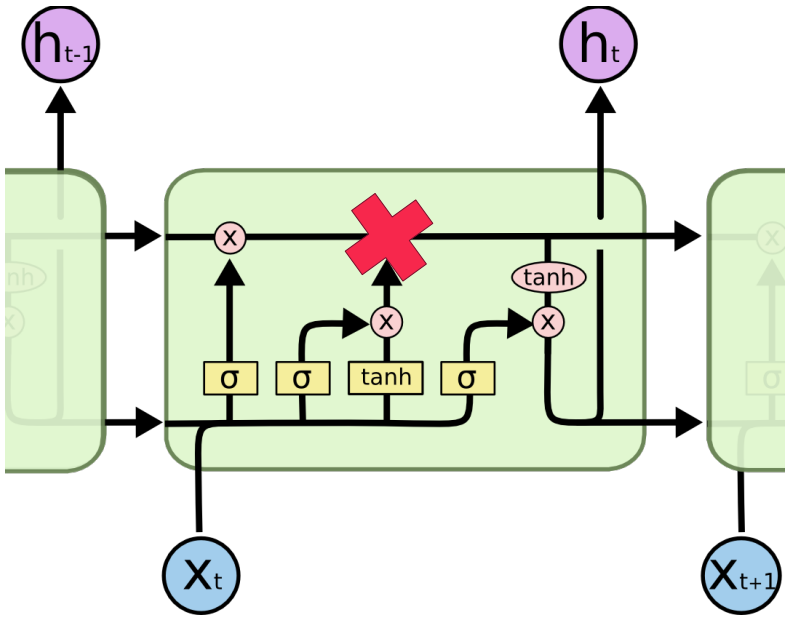
- No Input Gate (NIG)
- No Forget Gate (NFG)
- No Output Gate (NOG)
- No Input Activation Function (NIAF)
- No Output Activation Function (NOAF)
- No Peepholes (NP)
- Coupled Input and Forget Gate (CIFG) : GRU, $f_t = 1 - i_t$
- Full Gate Recurrence (FGR): Original LSTM paper



What's most important part of LSTM

Greff et al. explore

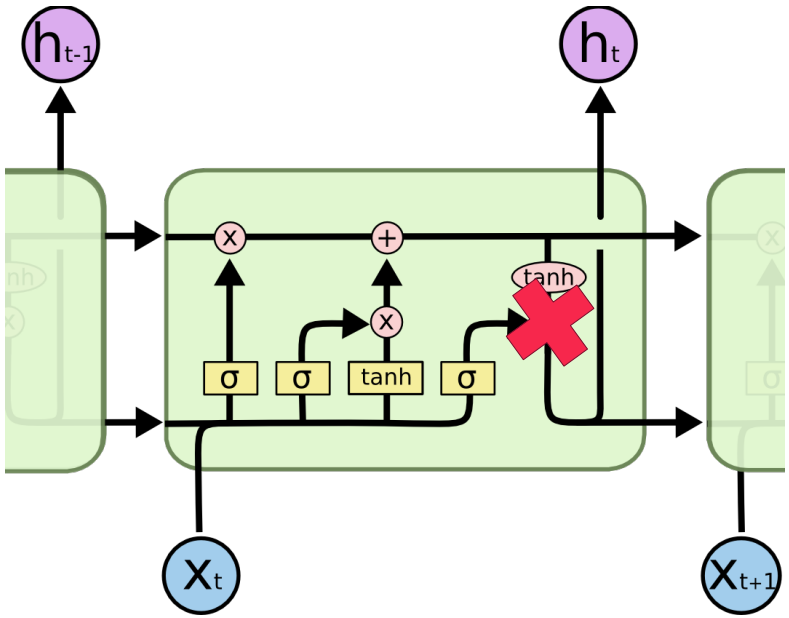
- No Input Gate (NIG)
- No Forget Gate (NFG)
- No Output Gate (NOG)
- No Input Activation Function (NIAF)
- No Output Activation Function (NOAF)
- No Peepholes (NP)
- Coupled Input and Forget Gate (CIFG) : GRU, $f_t = 1 - i_t$
- Full Gate Recurrence (FGR): Original LSTM paper



What's most important part of LSTM

Greff et al. explore

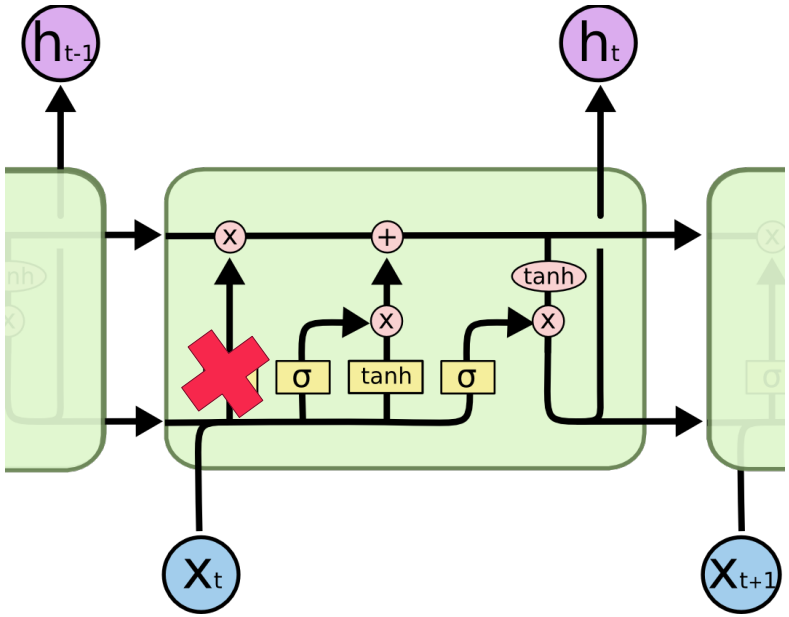
- No Input Gate (NIG)
- No Forget Gate (NFG)
- No Output Gate (NOG)
- No Input Activation Function (NIAF)
- No Output Activation Function (NOAF)
- No Peepholes (NP)
- Coupled Input and Forget Gate (CIFG) : GRU, $f_t = 1 - i_t$
- Full Gate Recurrence (FGR): Original LSTM paper



What's most important part of LSTM

Greff et al. explore

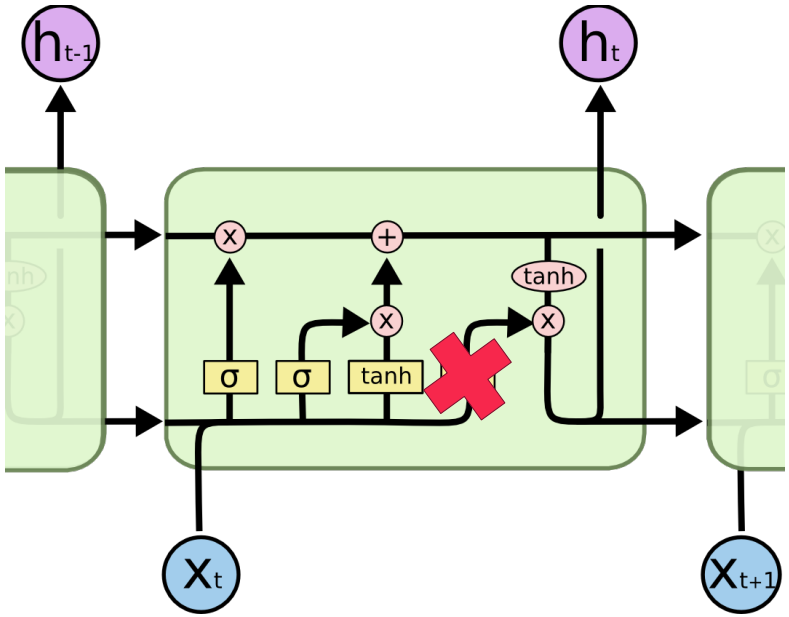
- No Input Gate (NIG)
- No Forget Gate (NFG)
- No Output Gate (NOG)
- No Input Activation Function (NIAF)
- No Output Activation Function (NOAF)
- No Peepholes (NP)
- Coupled Input and Forget Gate (CIFG) : GRU, $f_t = 1 - i_t$
- Full Gate Recurrence (FGR): Original LSTM paper



What's most important part of LSTM

Greff et al. explore

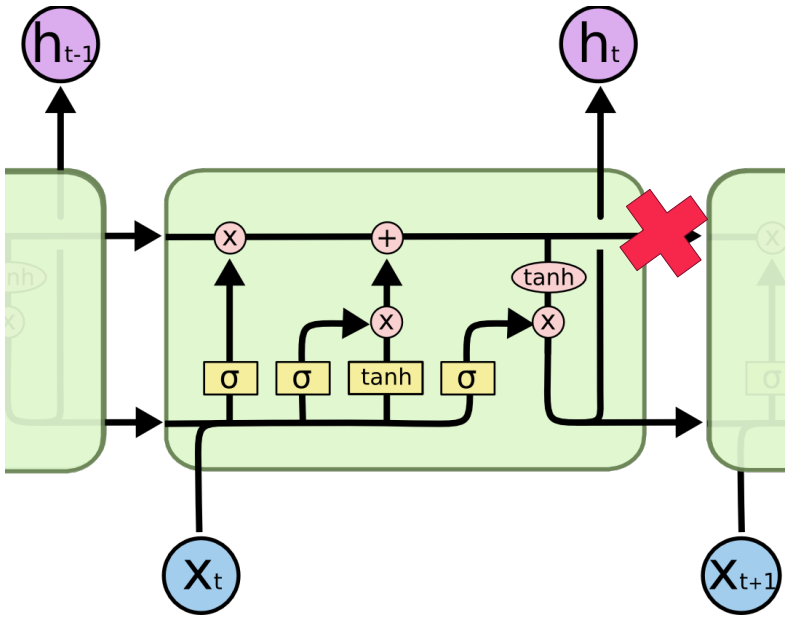
- No Input Gate (NIG)
- No Forget Gate (NFG)
- No Output Gate (NOG)
- No Input Activation Function (NIAF)
- No Output Activation Function (NOAF)
- No Peepholes (NP)
- Coupled Input and Forget Gate (CIFG) : GRU, $f_t = 1 - i_t$
- Full Gate Recurrence (FGR): Original LSTM paper



What's most important part of LSTM

Greff et al. explore

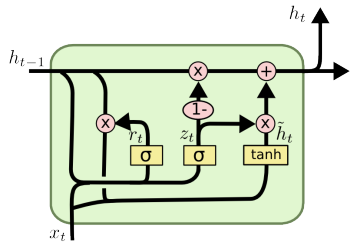
- No Input Gate (NIG)
- No Forget Gate (NFG)
- No Output Gate (NOG)
- No Input Activation Function (NIAF)
- No Output Activation Function (NOAF)
- No Peepholes (NP)
- Coupled Input and Forget Gate (CIFG) : GRU, $f_t = 1 - i_t$
- Full Gate Recurrence (FGR): Original LSTM paper



What's most important part of LSTM

Greff et al. explore

- No Input Gate (NIG)
- No Forget Gate (NFG)
- No Output Gate (NOG)
- No Input Activation Function (NIAF)
- No Output Activation Function (NOAF)
- No Peepholes (NP)
- Coupled Input and Forget Gate (CIFG) : GRU, $f_t = 1 - i_t$
- Full Gate Recurrence (FGR): Original LSTM paper



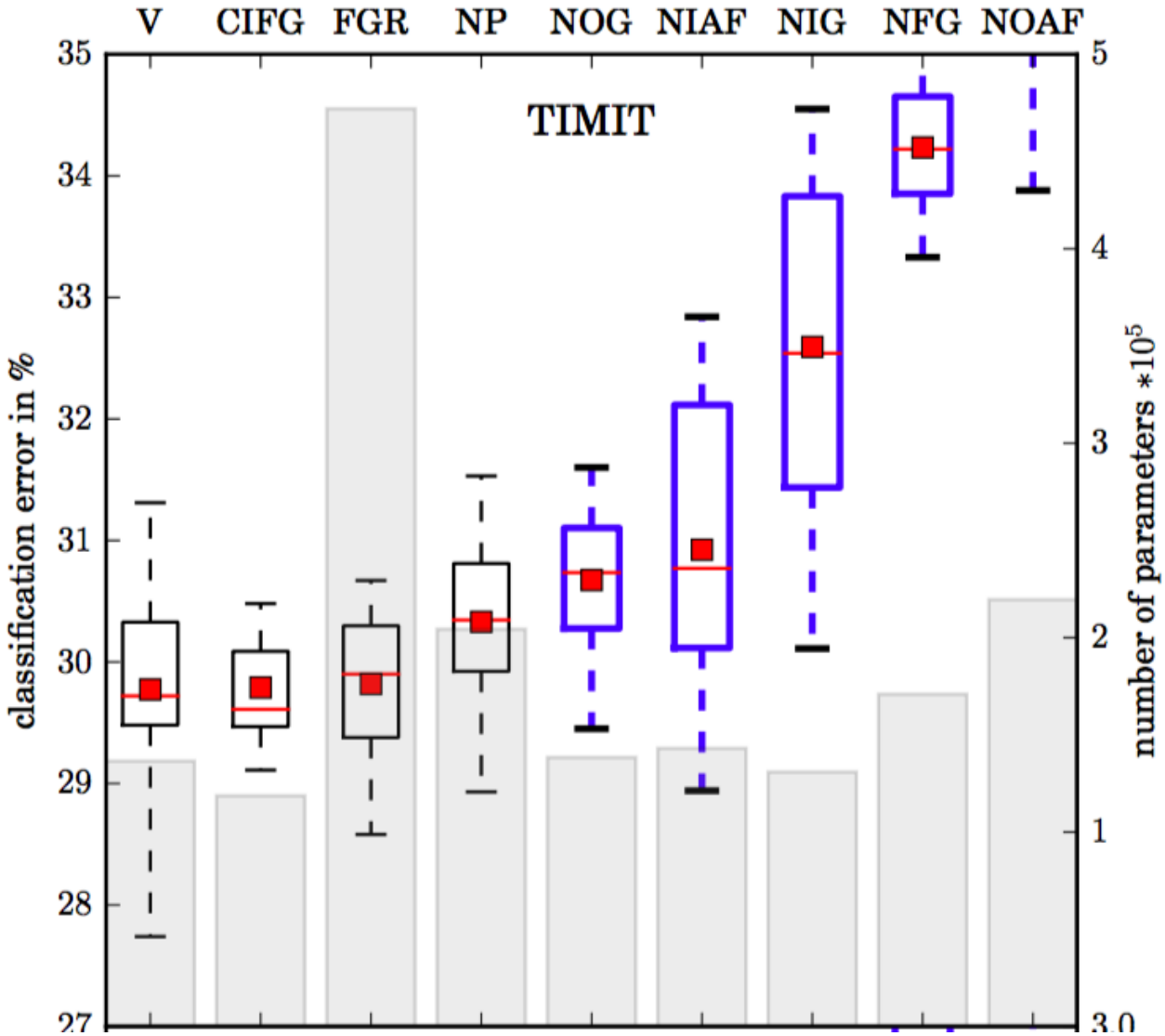
$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$
$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$
$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

What's most important part of LSTM

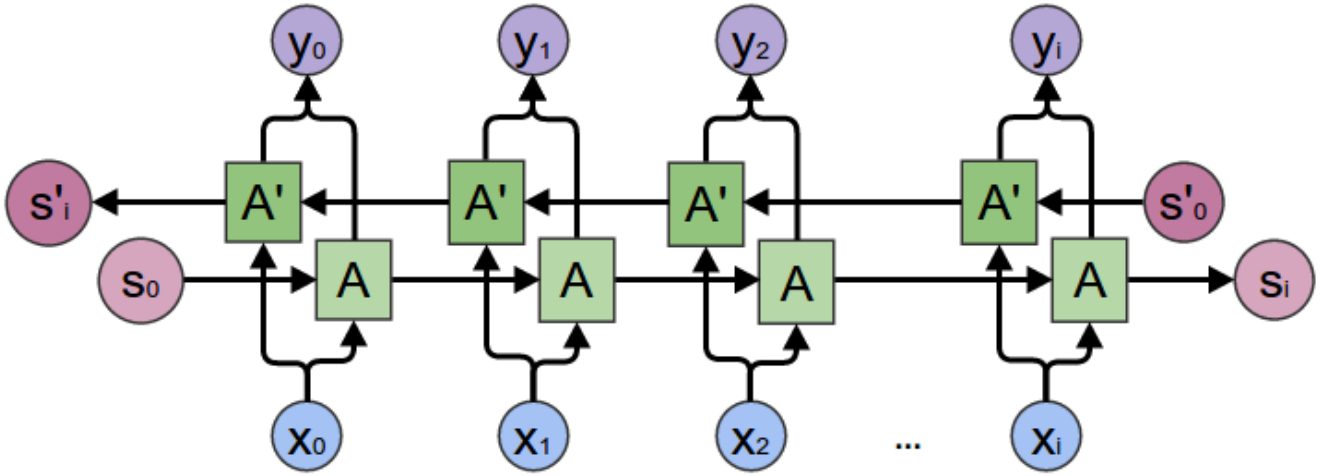
Greff et al. explore

- No Input Gate (NIG)
- No Forget Gate (NFG)
- No Output Gate (NOG)
- No Input Activation Function (NIAF)
- No Output Activation Function (NOAF)
- No Peepholes (NP)
- Coupled Input and Forget Gate (CIFG) : GRU, $f_t = 1 - i_t$
- Full Gate Recurrence (FGR): Original LSTM paper

What's most important part of LSTM



Bi-directional LSTMs



Simple extension, often slightly improve performance (but don't always make sense for task)

Comparing architechtures

- GRUs seem competitive
- LSTM seems to be good tradeoff
- Bi-directional often offers slight improvement