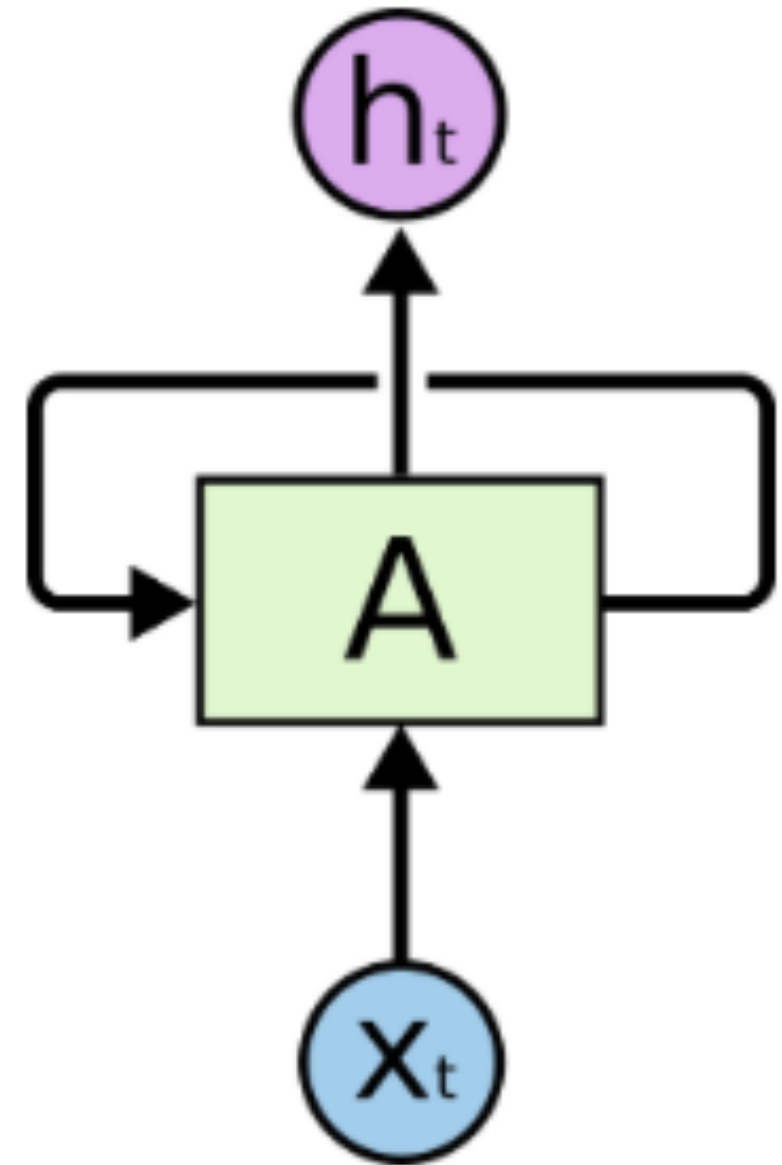


Введение в RNN и LSTM

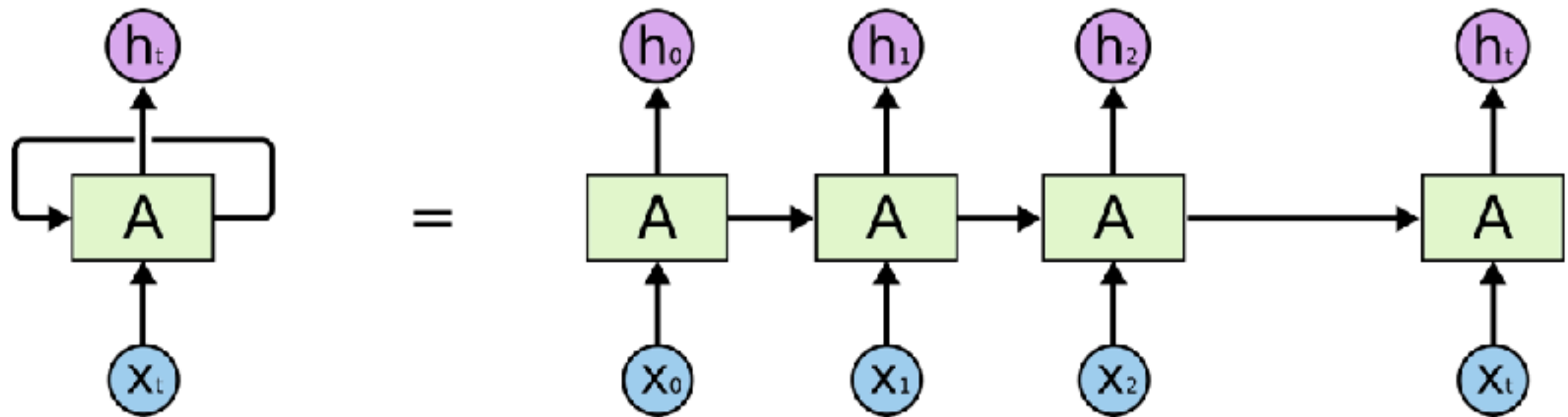
Студент БПМИ172, Валитов Эльдар

RNN. Описание модели

Получая на вход X_t , нейронная сеть A выдает на выходе h_t

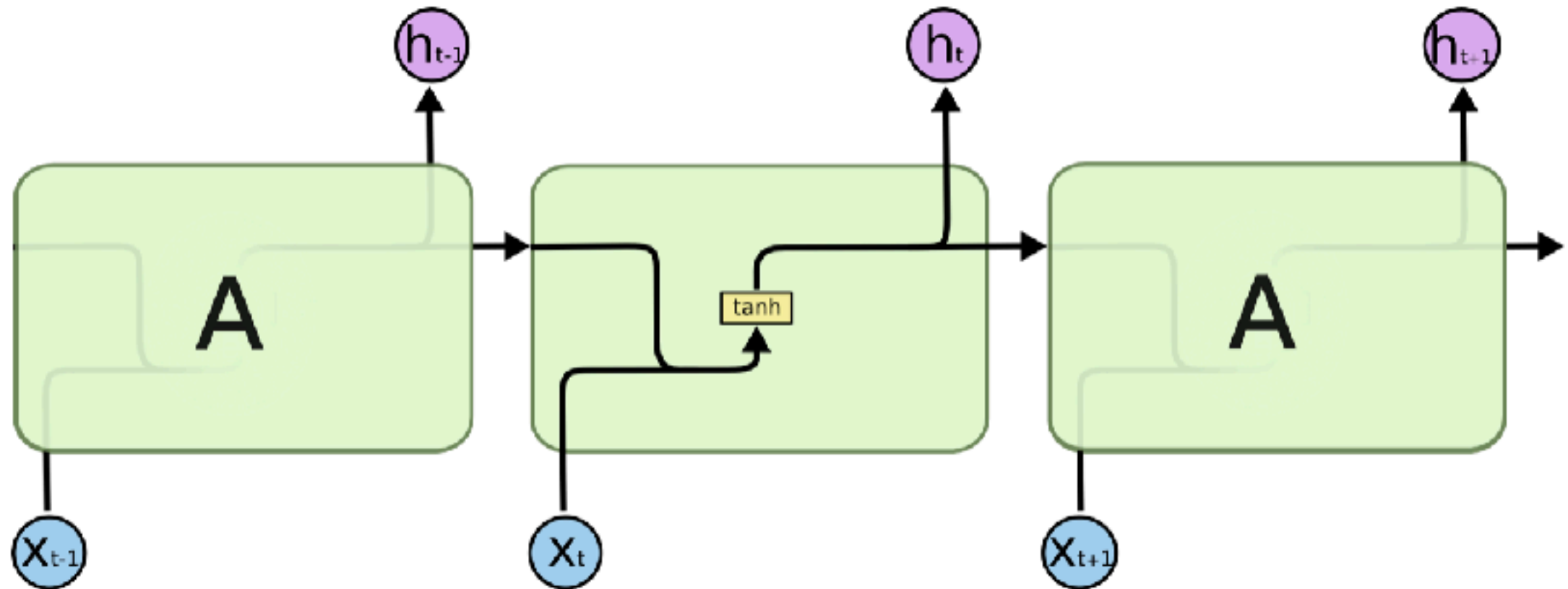


RNN. Описание модели



Развернутая RNN.

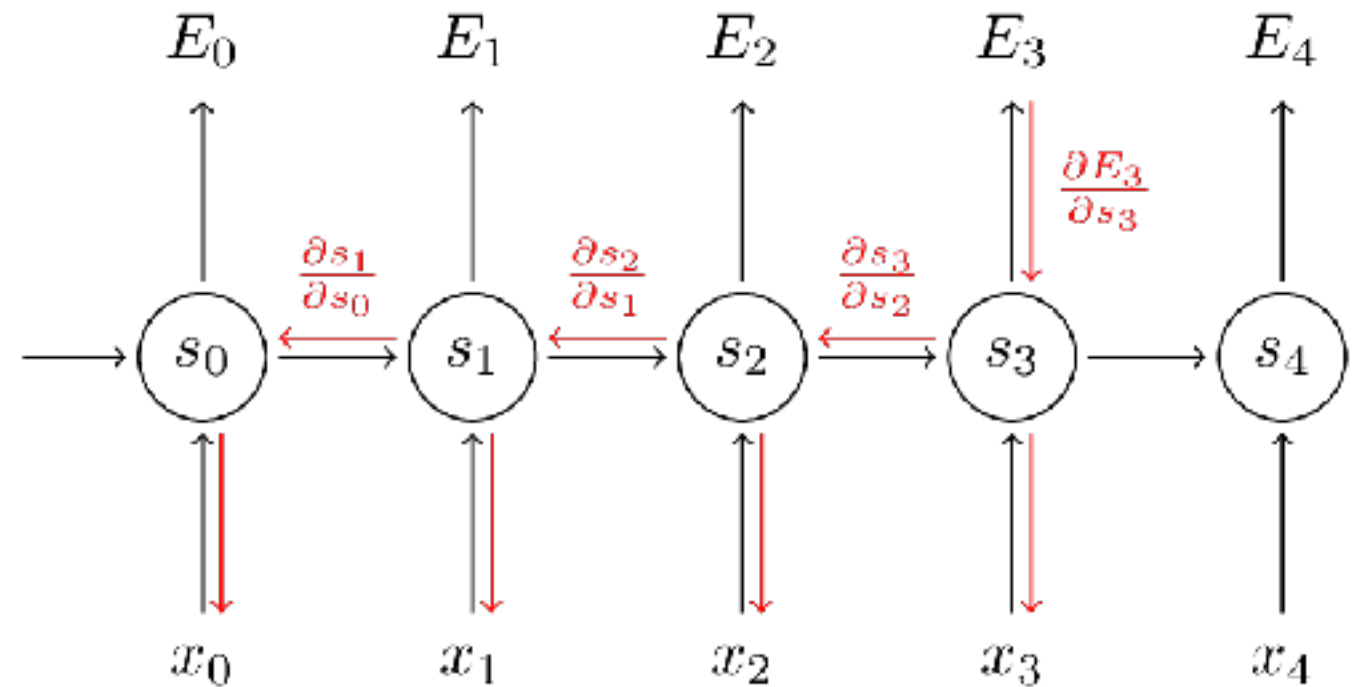
RNN. Описание модели



Повторяющийся модуль в RNN

RNN. Обучение модели

RNN обучаются с помощью вариации Backpropagation - "Backpropagation through time" (BPTT)



Плюс и минусы RNN

Плюс

- Последовательная структура позволяет предыдущим входным данным влиять на новые

Минусы

- Сложность вычислений
- Взрывающиеся/затухающие градиенты (Vanishing/Exploding gradients problem)

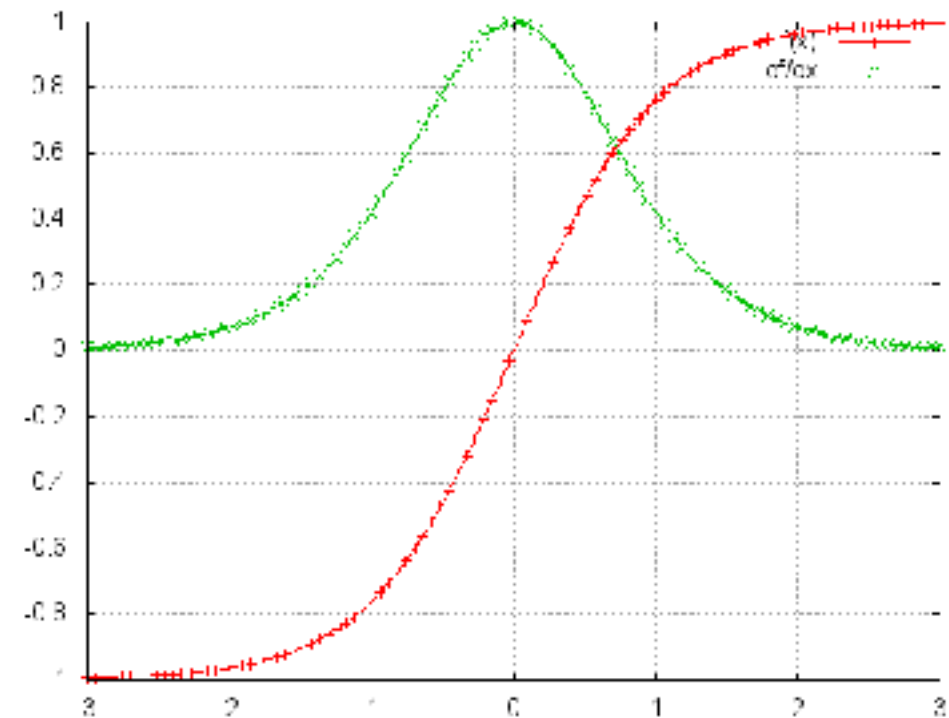
Мотивация к созданию LSTM

$$\prod_{t=2}^k \tanh'(W_c C_{t-1} + W_x X_t) \cdot W_c \rightarrow 0, \text{ so } \frac{\partial E_k}{\partial W} \rightarrow 0$$

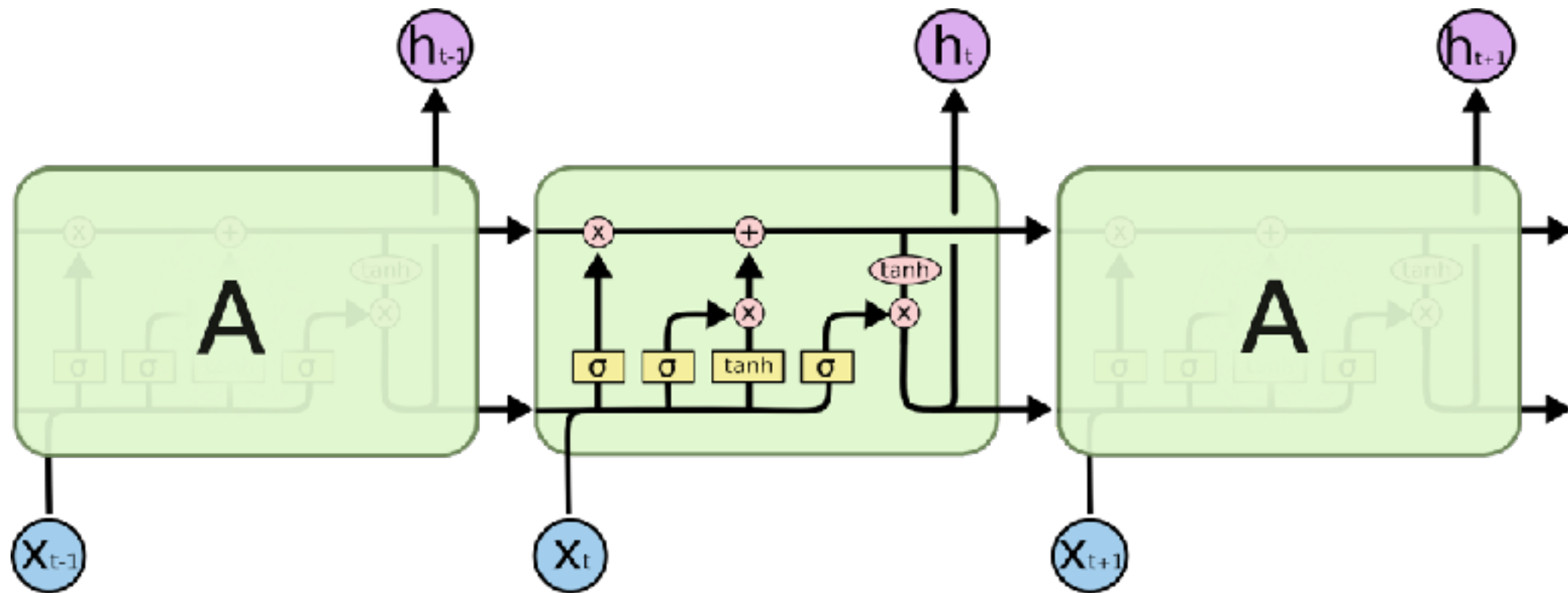
Проблема появляется из-за производных tanh и сигмоиды равных нулю

Взрывающиеся градиенты страшно, но
затухающие страшнее - из-за них
появляются NaN

$$W \leftarrow W - \alpha \frac{\partial E_k}{\partial W} \approx W$$

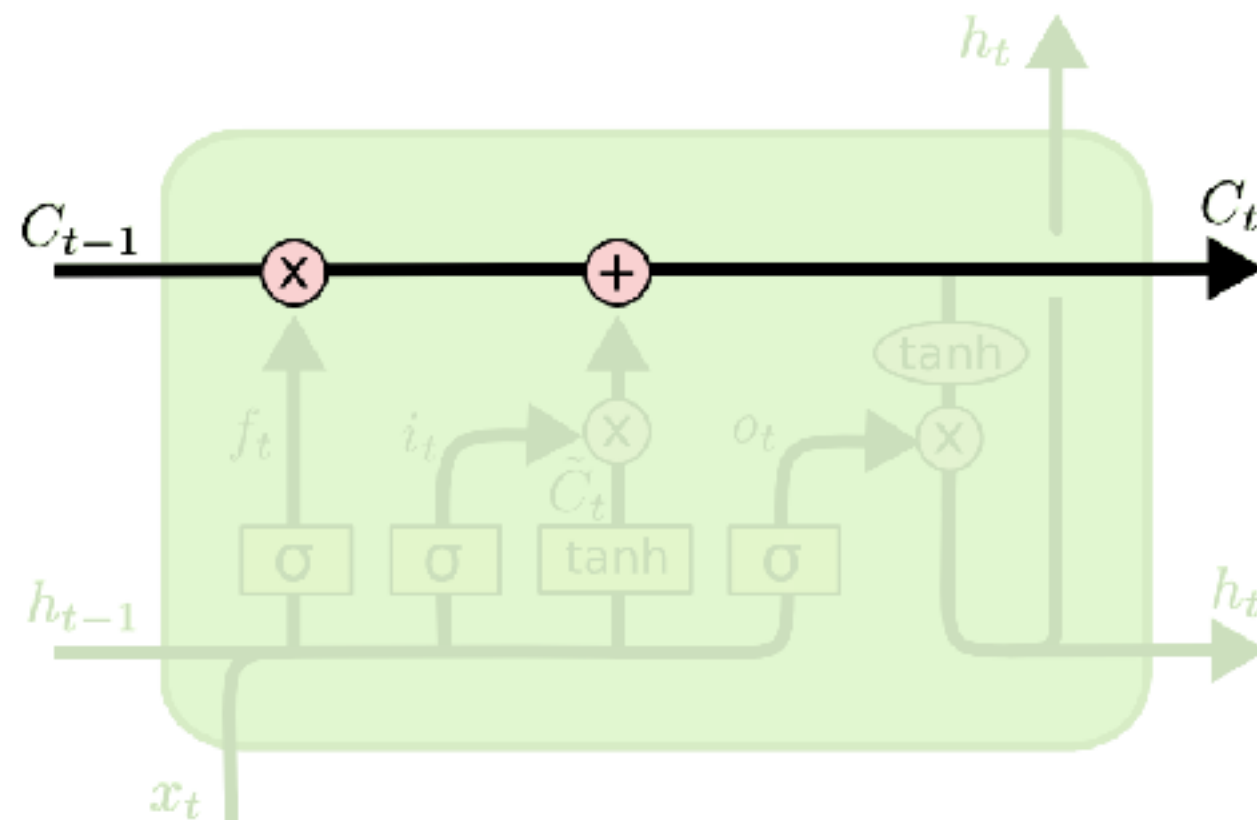


LSTM. Описание модели



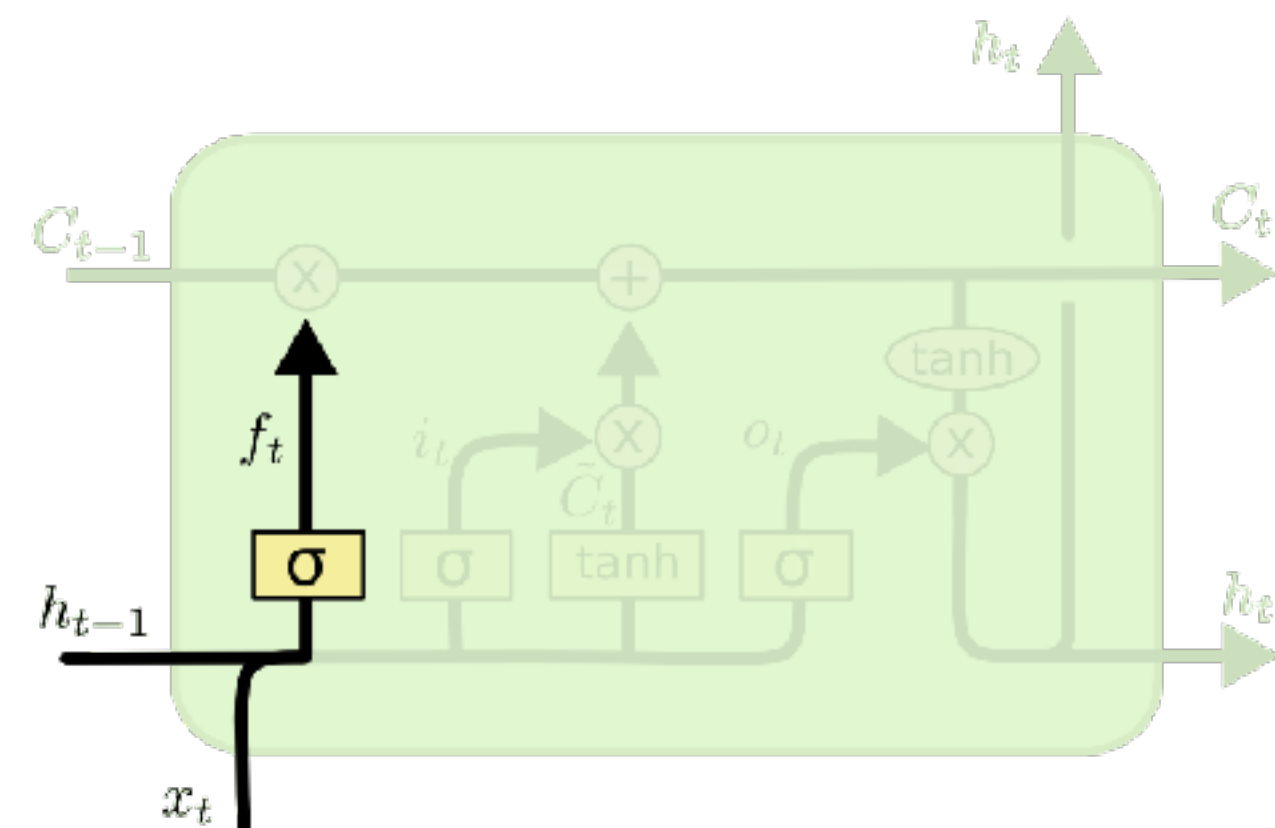
Повторяющийся модуль в LSTM

LSTM. Описание модели



Основной слой данных

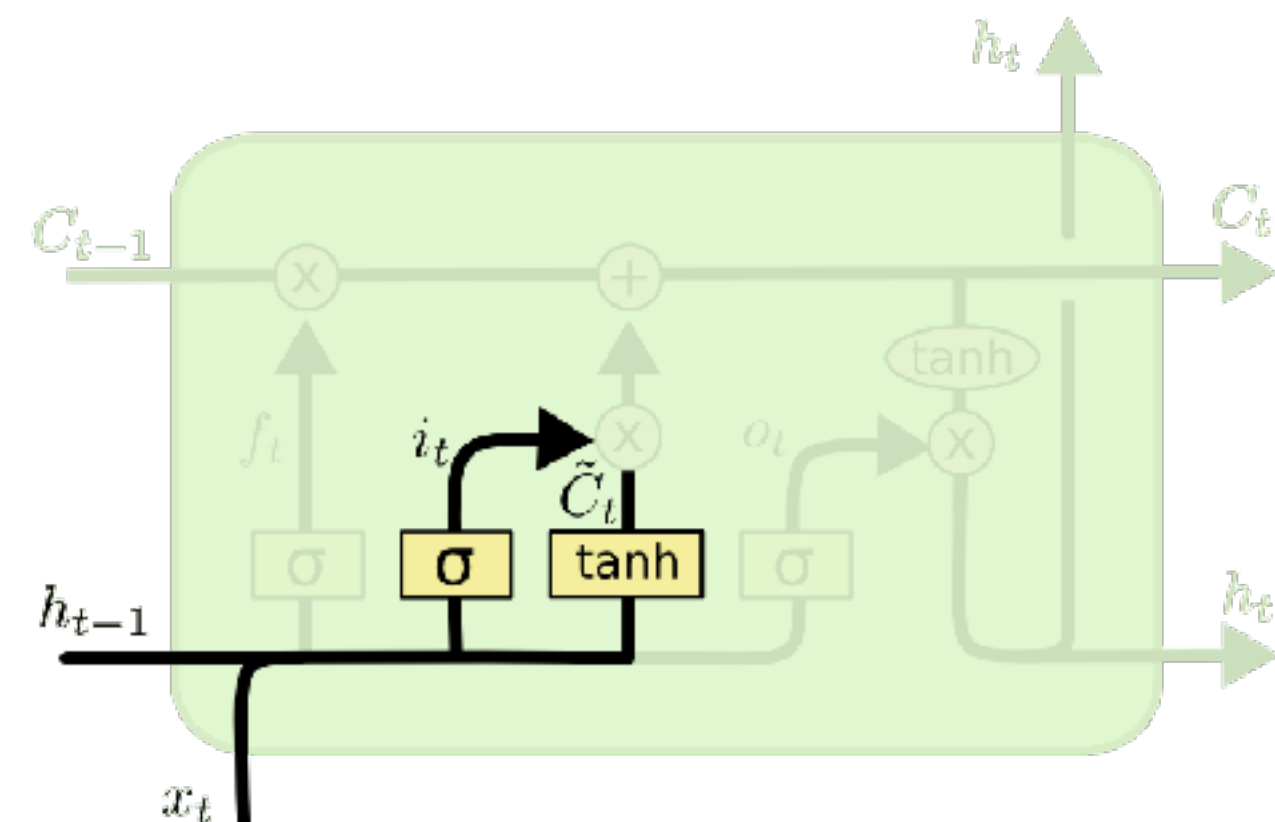
LSTM. Описание модели



$$f_t = \sigma (W_f \cdot [h_{t-1}, x_t] + b_f)$$

"forget gate layer" - слой забвения

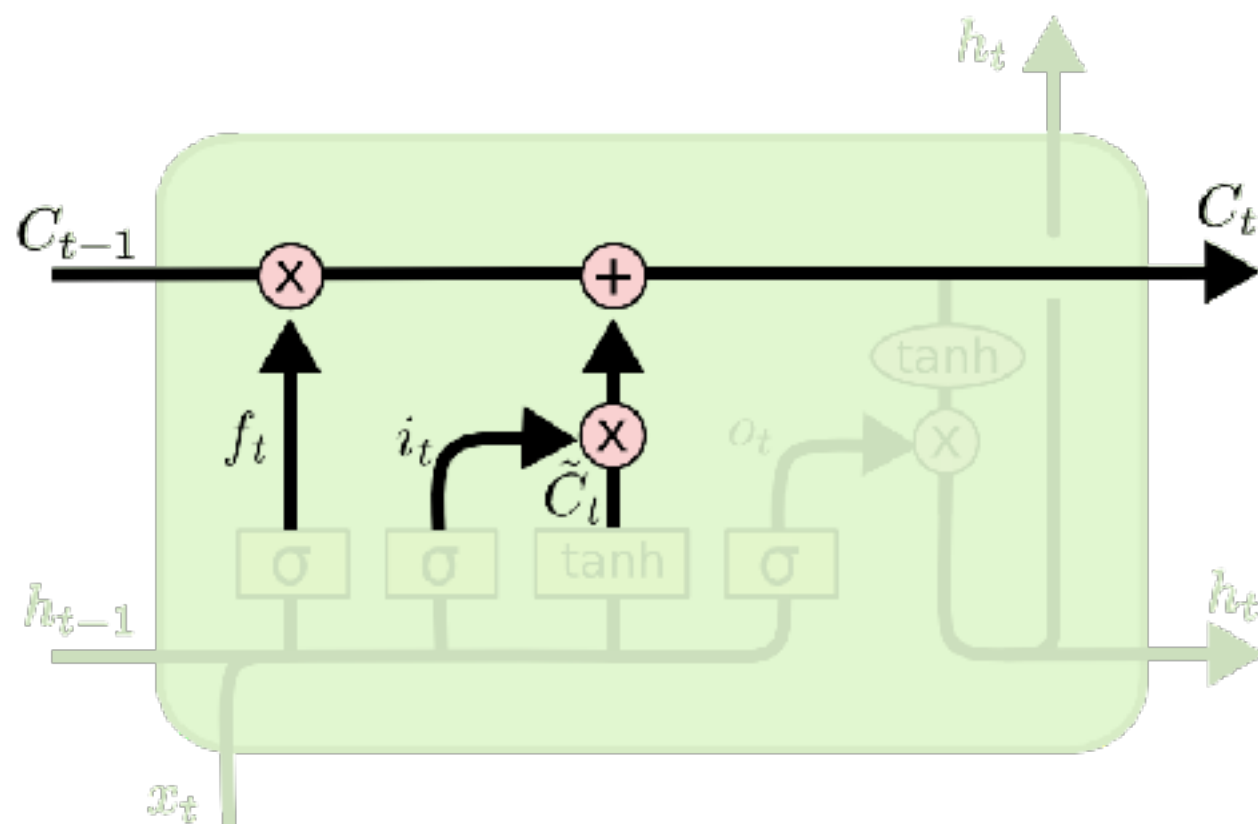
LSTM. Описание модели



$$i_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

"input gate layer" - входной слой

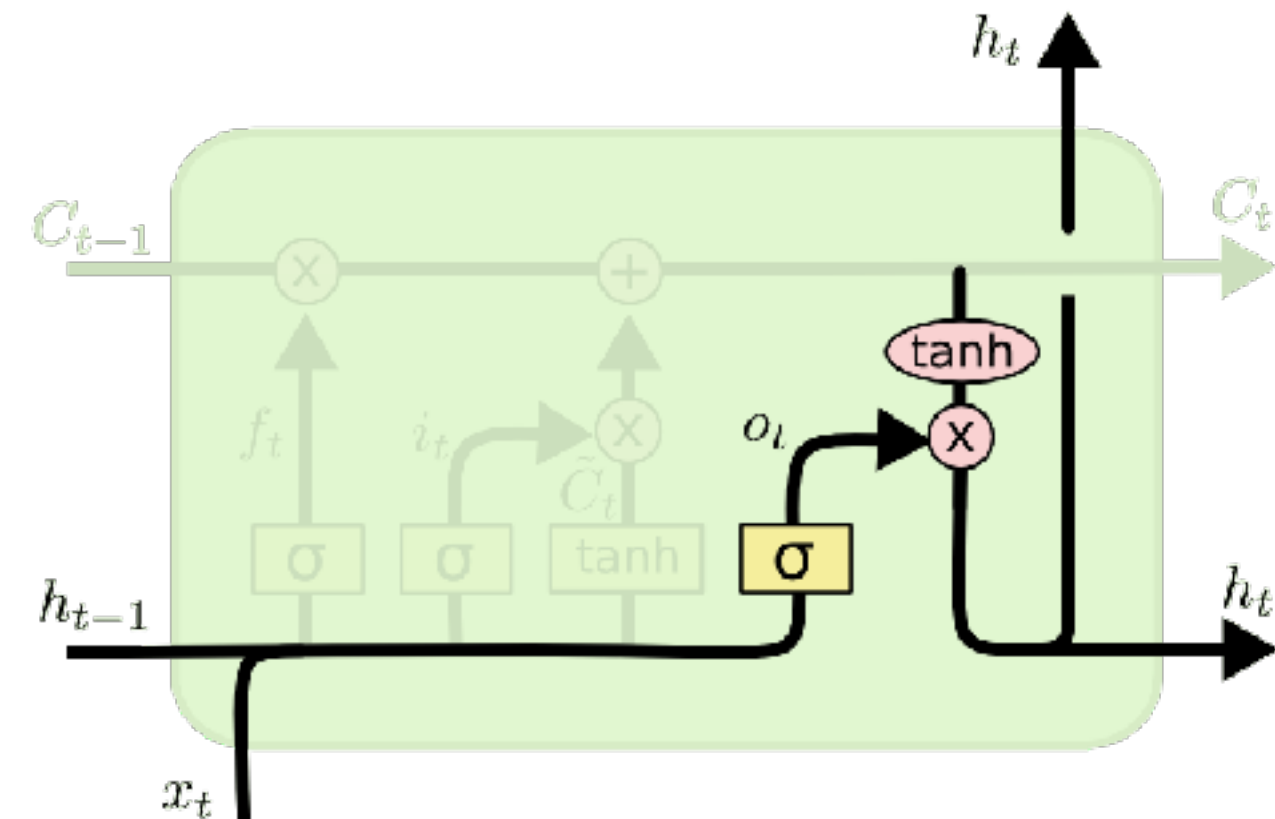
LSTM. Описание модели



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Обновление значений

LSTM. Описание модели



$$o_t = \sigma (W_o [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh (C_t)$$

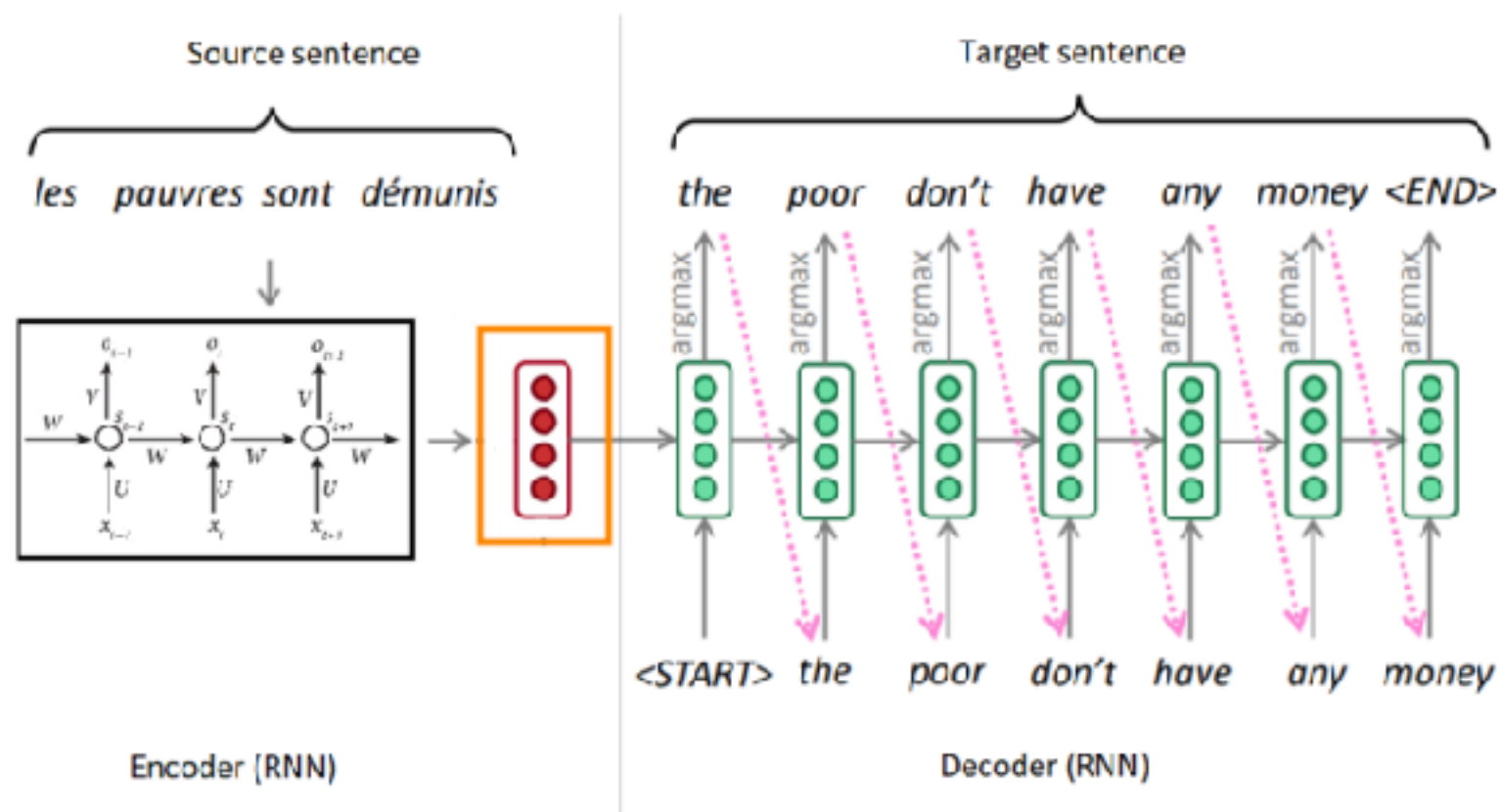
"Output gate layer" - выходной слой

$$\frac{\partial E_k}{\partial W} \approx \frac{\partial E_k}{\partial H_k} \frac{\partial H_k}{\partial C_k} \left(\prod_{t=2}^k \sigma(W_f \cdot [H_{t-1}, X_t]) \right) \frac{\partial C_1}{\partial W} \nrightarrow 0$$

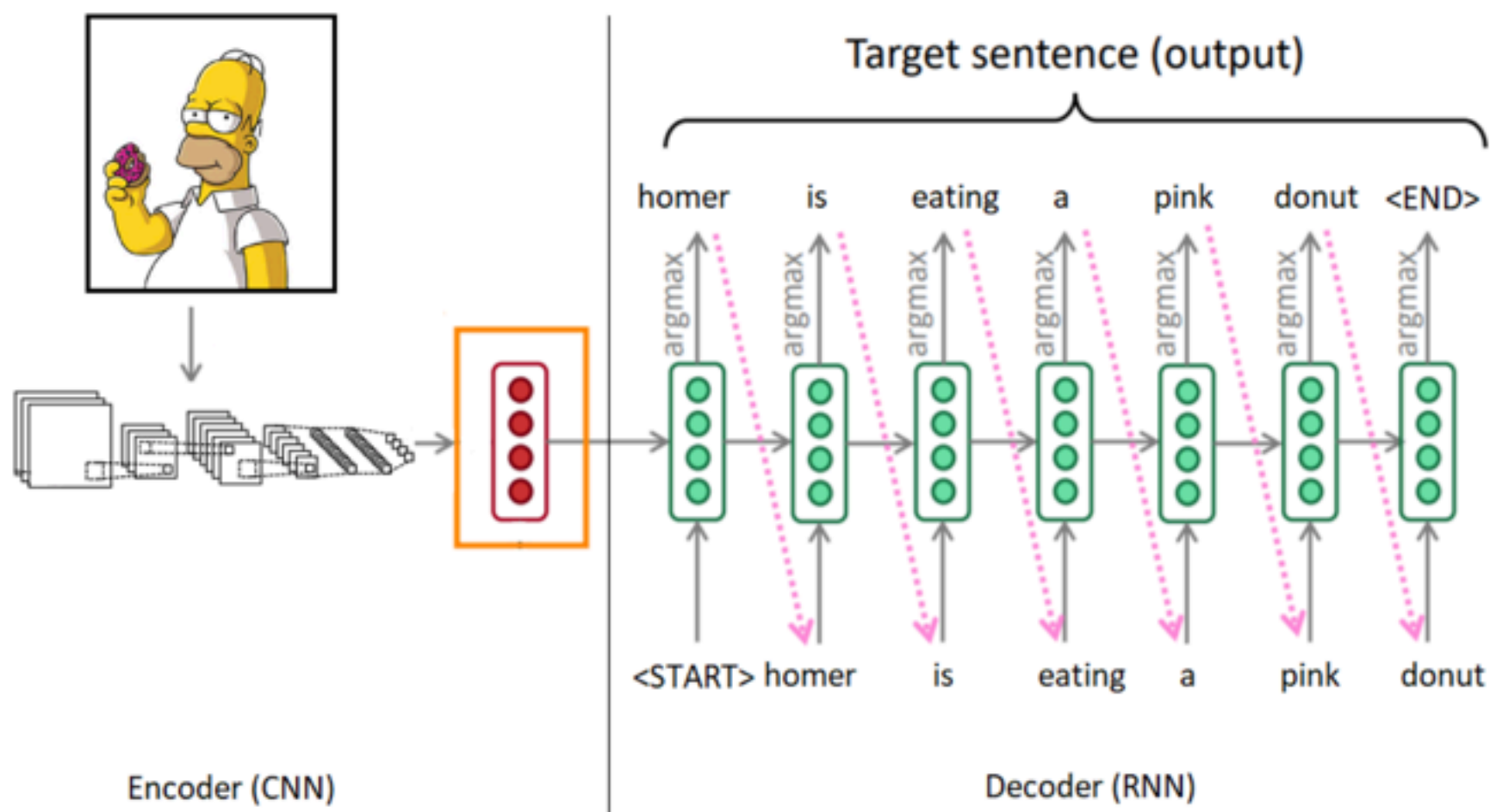
LSTM или RNN?

- В большинстве случаев LSTM работает лучше, чем RNN
- Когда есть необходимость сэкономить ресурсы, то лучше использовать RNN

Машинный перевод



Описание изображения



Вопросы

- Опишите кратко модификацию LSTM над RNN.
- Когда имеет смысл использовать RNN вместо LSTM?
- Опишите проблему затухающего градиента.

ИСТОЧНИКИ

- <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- <http://www.bioinf.jku.at/publications/older/2604.pdf>
- <https://arxiv.org/pdf/1808.03314.pdf>
- <http://proceedings.mlr.press/v28/pascanu13.pdf>
- <http://www.wildml.com/2015/10/recurrent-neural-networks-tutorial-part-3-backpropagation-through-time-and-vanishing-gradients/>
- <https://medium.com/datadriveninvestor/how-do-lstm-networks-solve-the-problem-of-vanishing-gradients-a6784971a577>