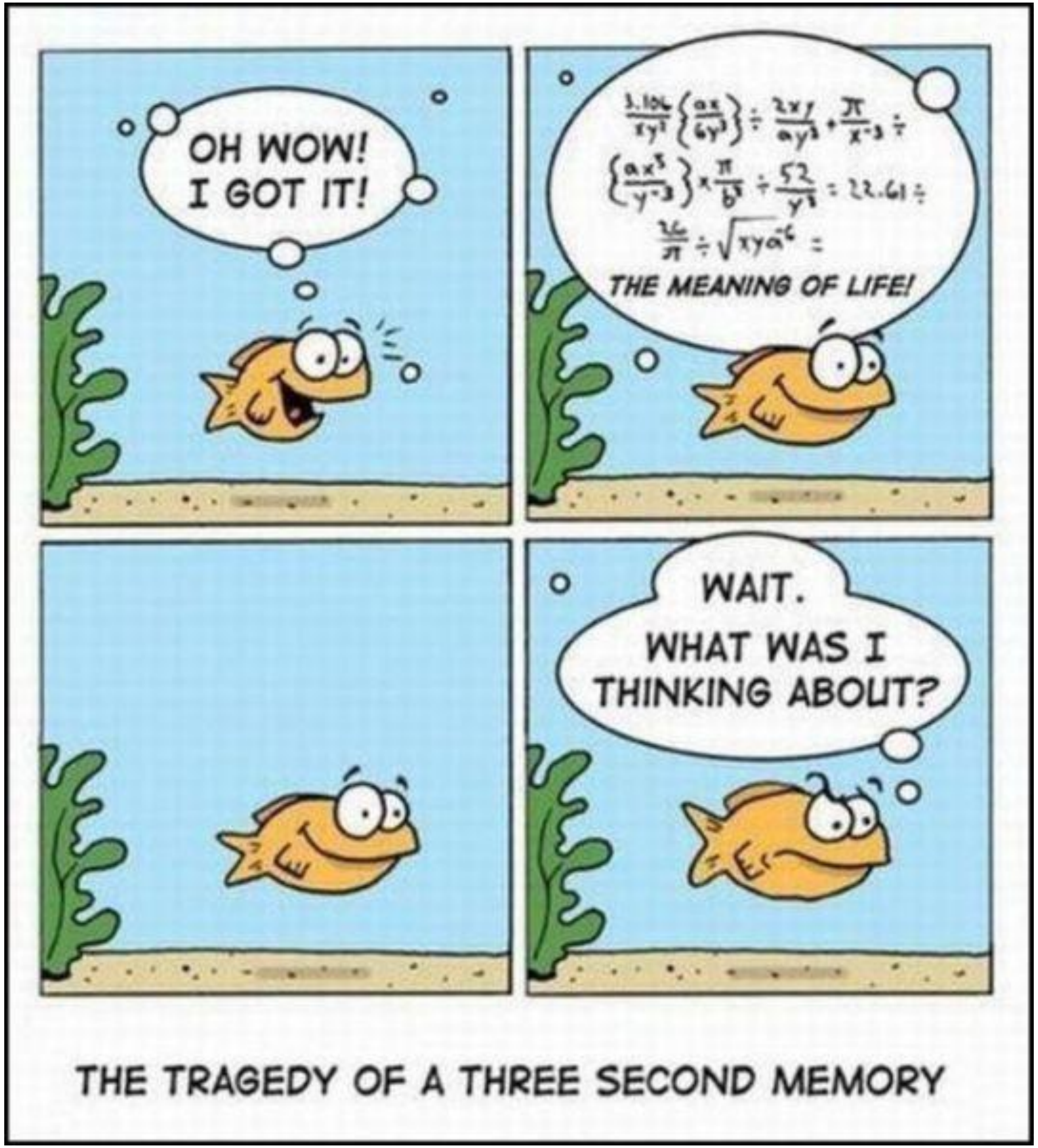
RNN + LSTM

Бондаренко Наталия, БПМИ171

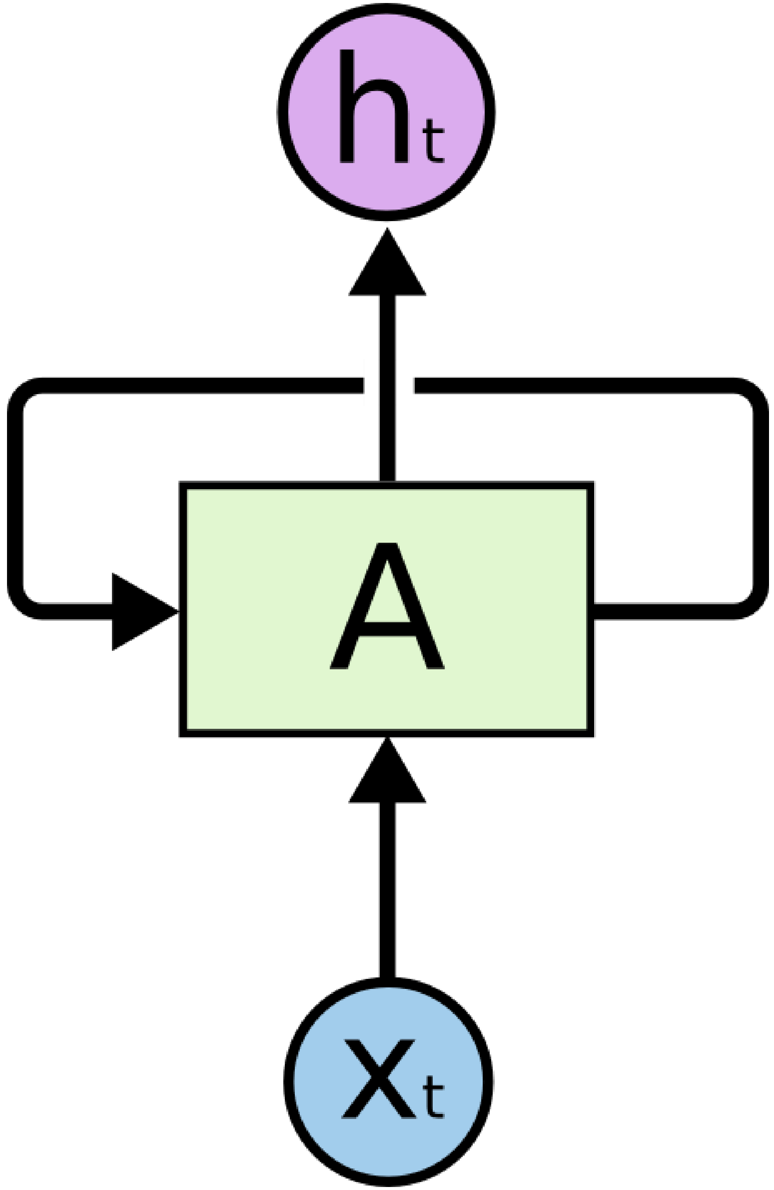
Проблема других архитектур



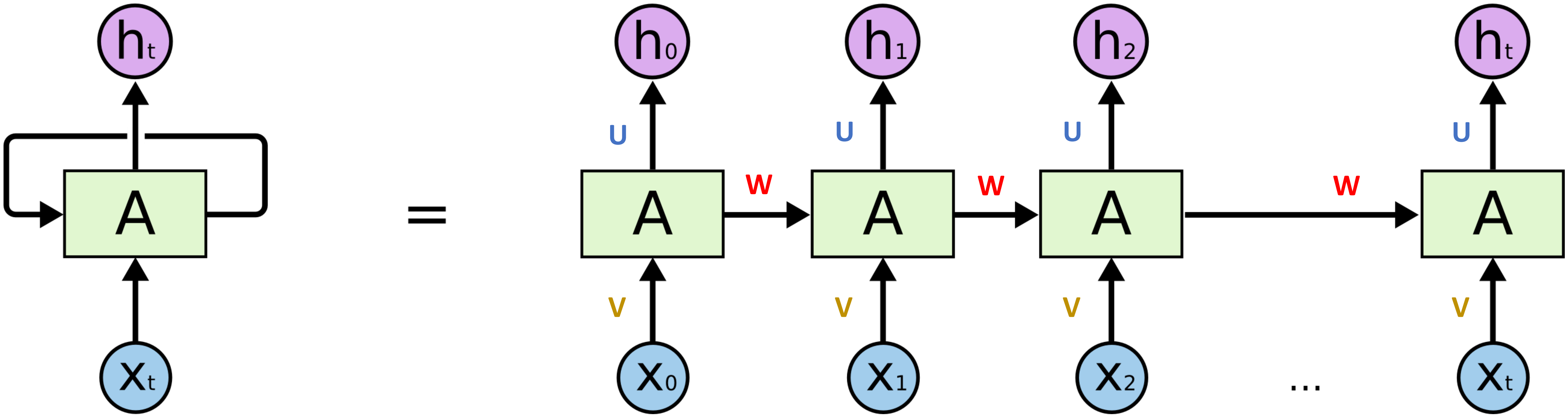
THE TRAGEDY OF A THREE SECOND MEMORY

RNN

- Содержат обратные связи
- Тьюринг-полны (реализуют любую вычислимую функцию)
- Обладают памятью

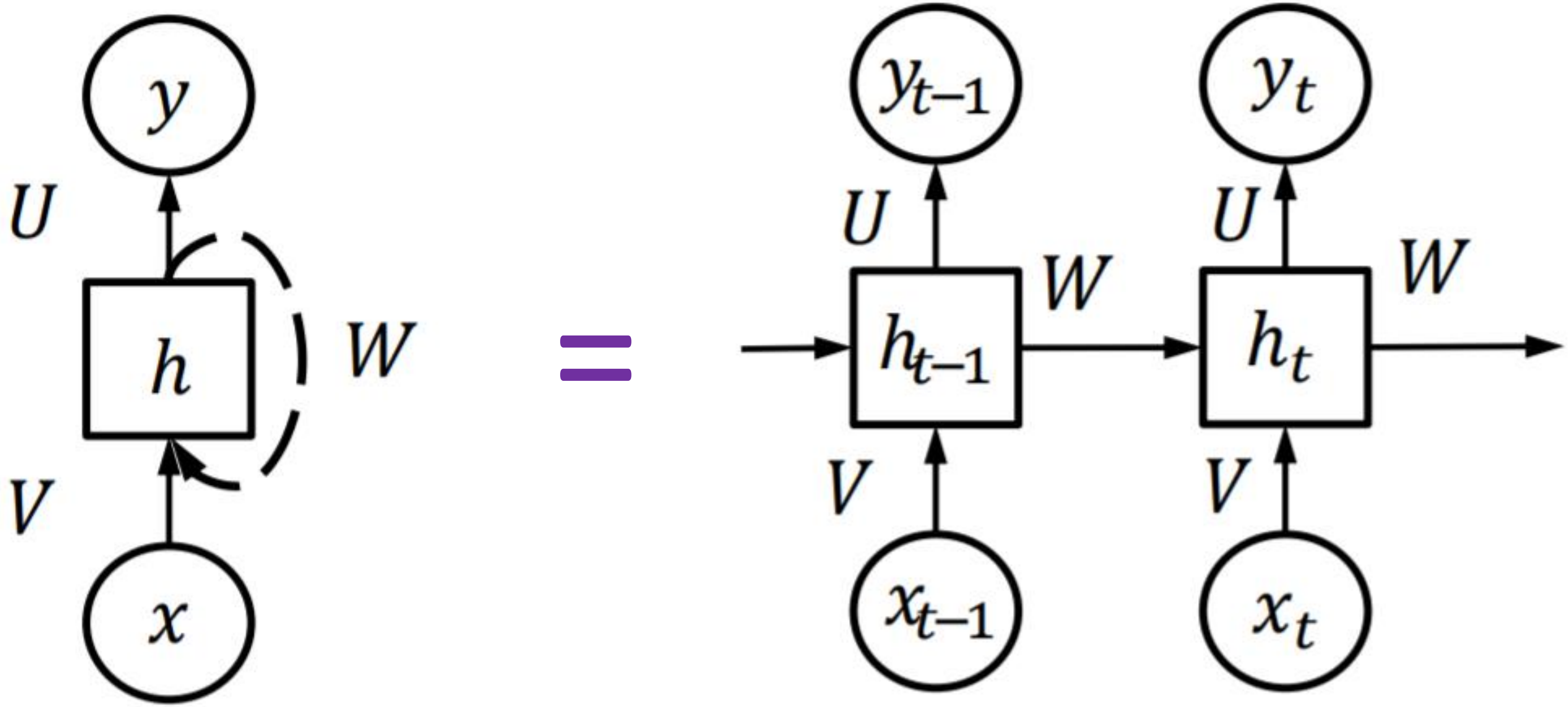


На самом деле



На самом деле

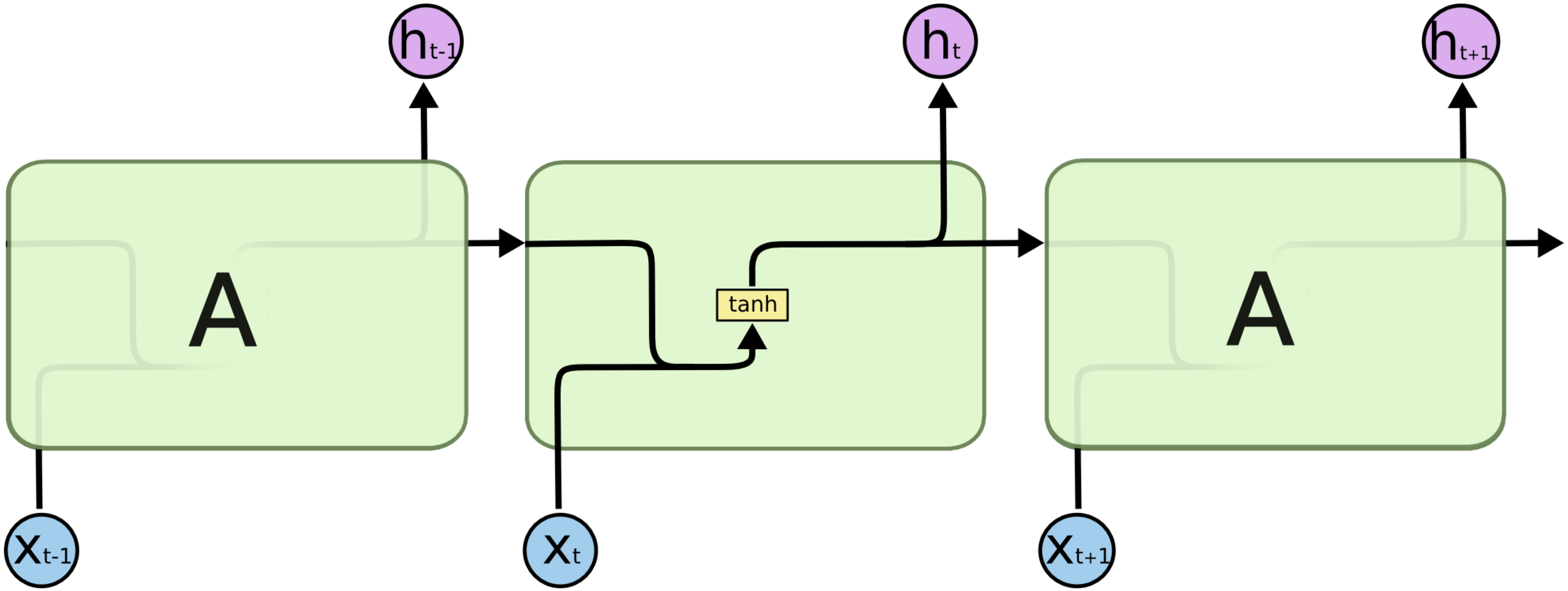
$$\boxed{\begin{gathered} h_t = g(Vx_t + Wh_{t-1} + b_h) \\ y_t = f(Uh_t + b_y) \end{gathered}}$$



RNN

Функция потерь: $$\boxed{L(y, a(x)) = \sum_t L(y_t, a(x_t))}$$

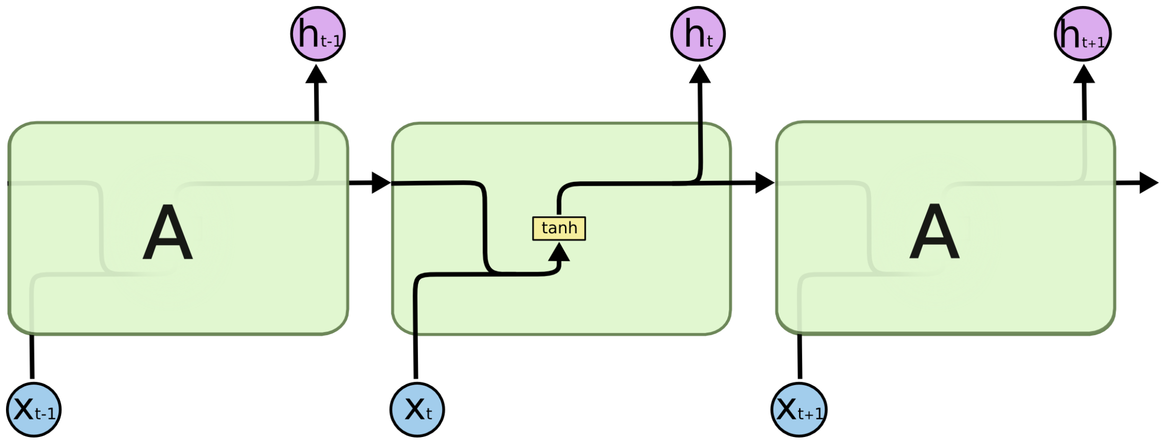
Слой стандартной RNN:



Backpropagation through time

Функция потерь:

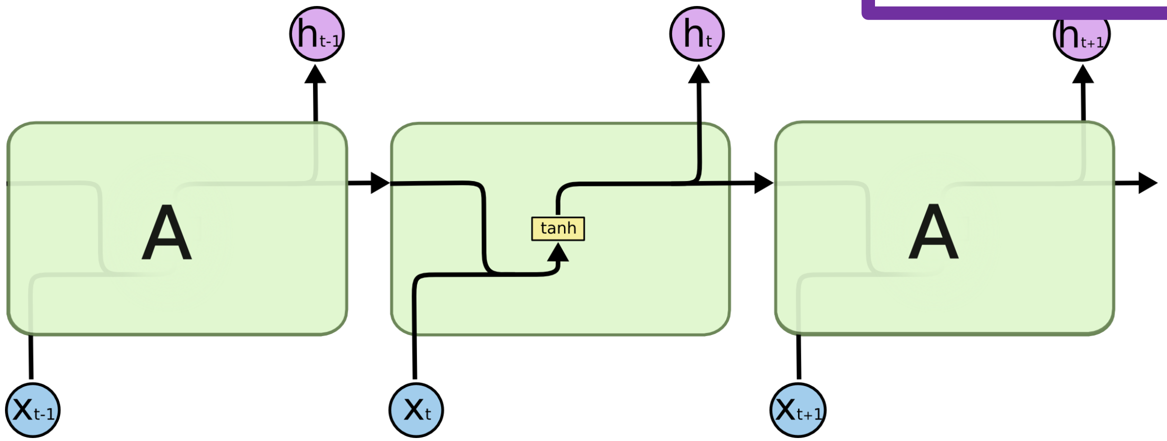
$$\boxed{L(y, a(x)) = \sum_t L(y_t, a(x_t))}$$



Backpropagation through time

Функция потерь: $$L(y, a(x)) = \sum_t L(y_t, a(x_t))$$

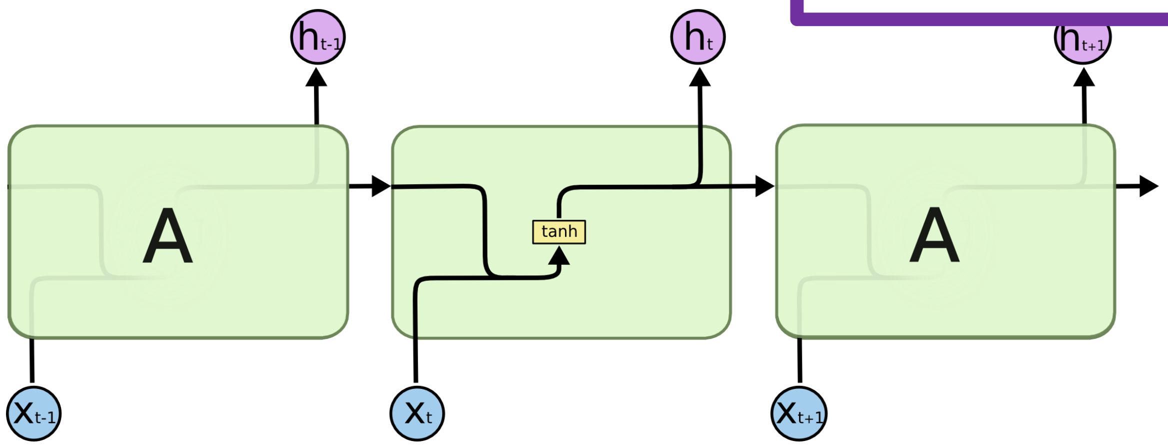
Ее производная: $$\frac{\partial L}{\partial W} = \sum_t \frac{\partial L_t}{\partial W_t}$$



Backpropagation through time

Функция потерь: $$\boxed{L(y, a(x)) = \sum_t L(y_t, a(x_t))}$$

Ее производная: $$\boxed{\frac{\partial L}{\partial W} = \sum_{i \le t} \frac{\partial L_i}{\partial W_i} = \sum_{i \le t} \frac{\partial L}{\partial h_t} \cdot \frac{\partial h_t}{\partial h_i} \cdot \frac{\partial h_i}{\partial W}}$$



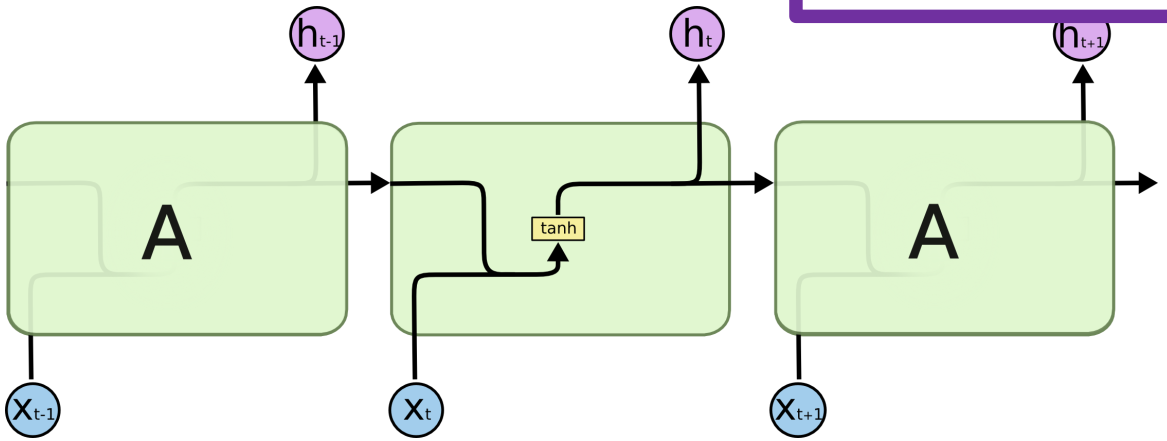
Backpropagation through time

Функция потерь:

$$L(y, a(x)) = \sum_t L(y_t, a(x_t))$$

Ее производная:

$$\frac{\partial L}{\partial W} = \sum_{i \leq t} \frac{\partial L_i}{\partial W_i} = \sum_{i \leq t} \frac{\partial L}{\partial h_t} \cdot \boxed{\frac{\partial h_t}{\partial h_i}} \cdot \frac{\partial h_i}{\partial W}$$



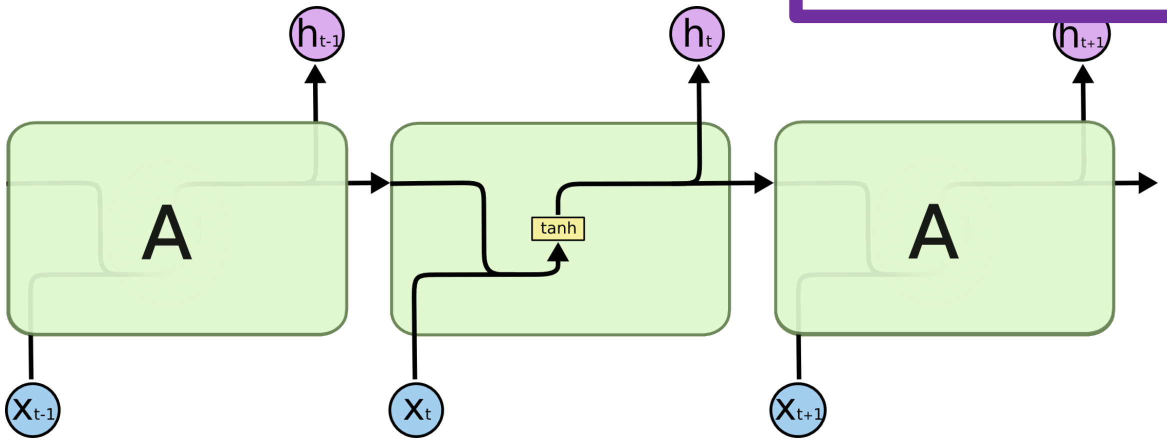
Backpropagation through time

Функция потерь:

$$\boxed{L(y, a(x)) = \sum_t L(y_t, a(x_t))}$$

Ее производная:

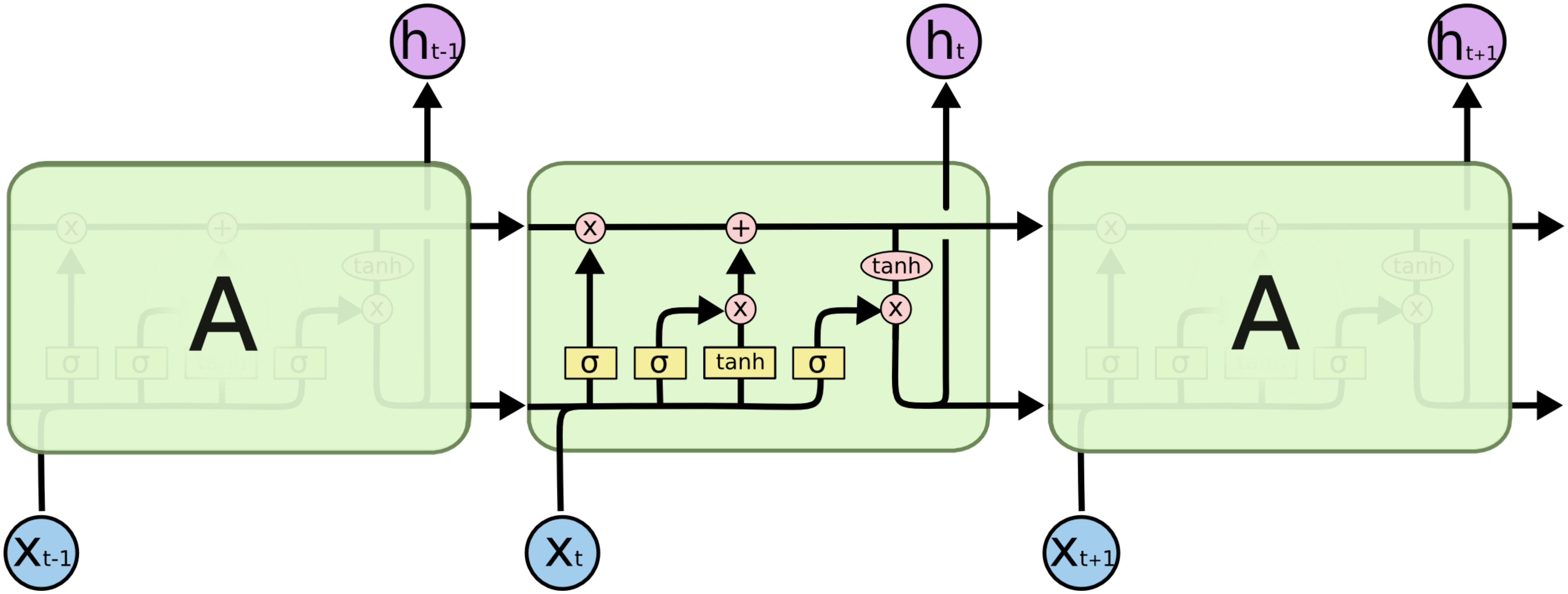
$$\boxed{\frac{\partial L}{\partial W} = \sum_{i \le t} \frac{\partial L_i}{\partial W_i} = \sum_{i \le t} \frac{\partial L}{\partial h_t} \cdot \boxed{\frac{\partial h_t}{\partial h_i}} \cdot \frac{\partial h_i}{\partial W}}$$



$$\boxed{\frac{\partial h_t}{\partial h_{t-1}} \cdot \frac{\partial h_{t-1}}{\partial h_{t-2}} \cdot \ldots \cdot \frac{\partial h_{i+1}}{\partial h_i}}$$

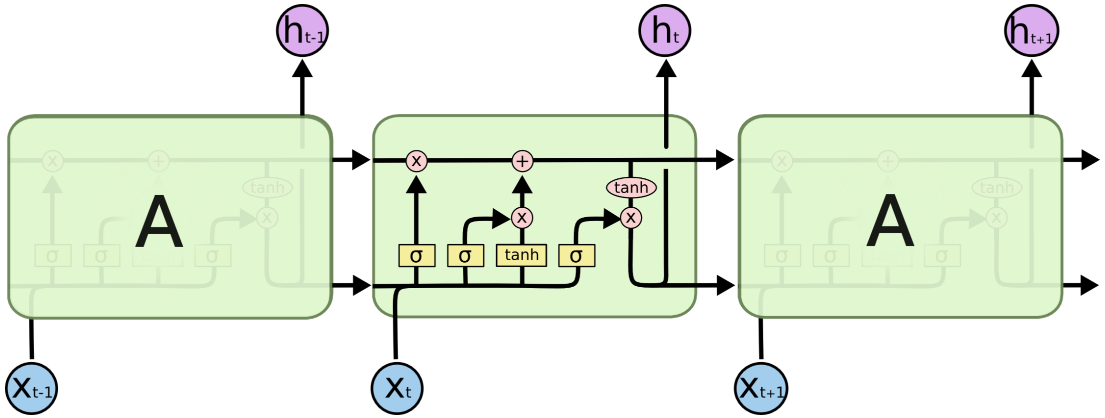
LSTM

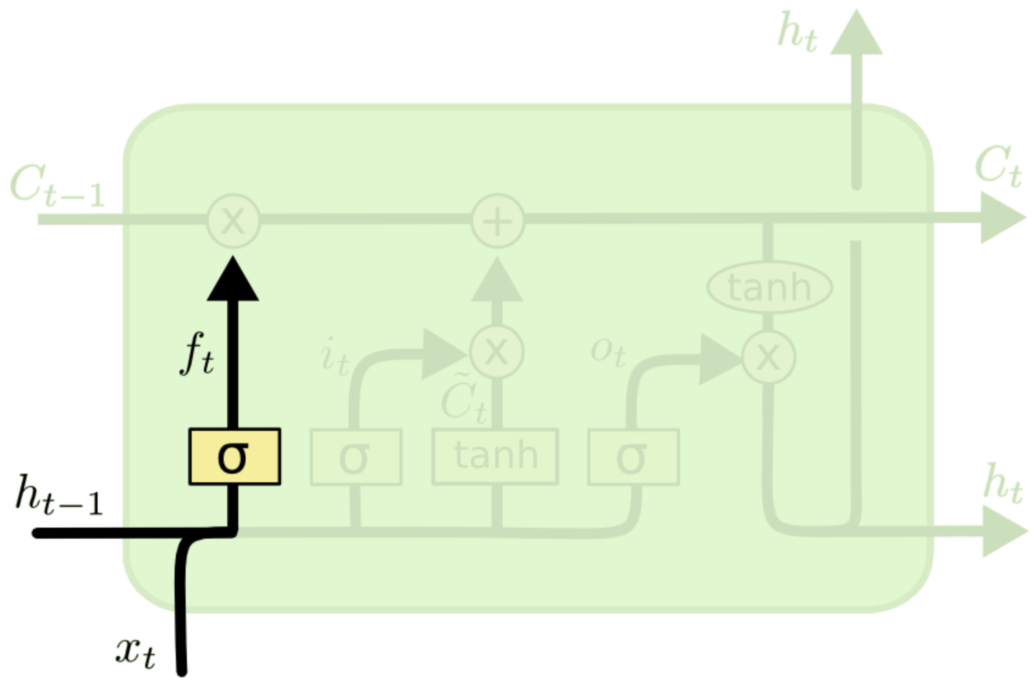
Как выглядит LSTM:



LSTM

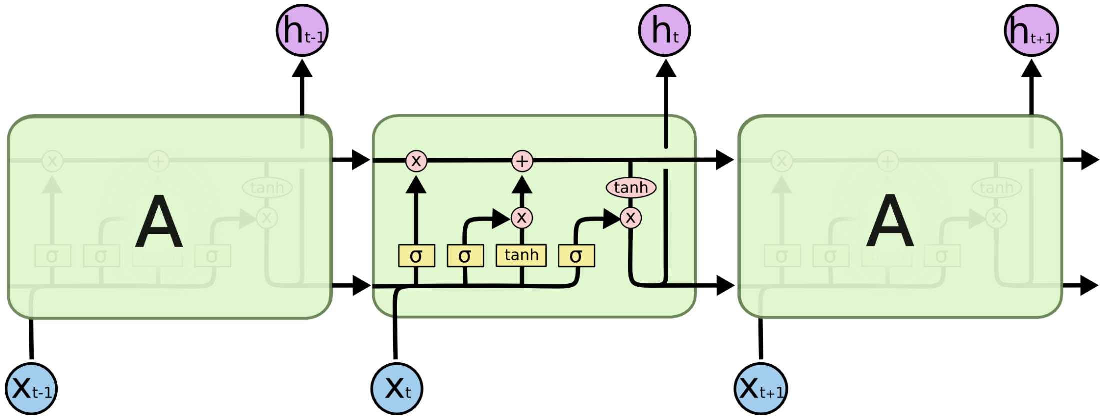
Фильтр забывания:



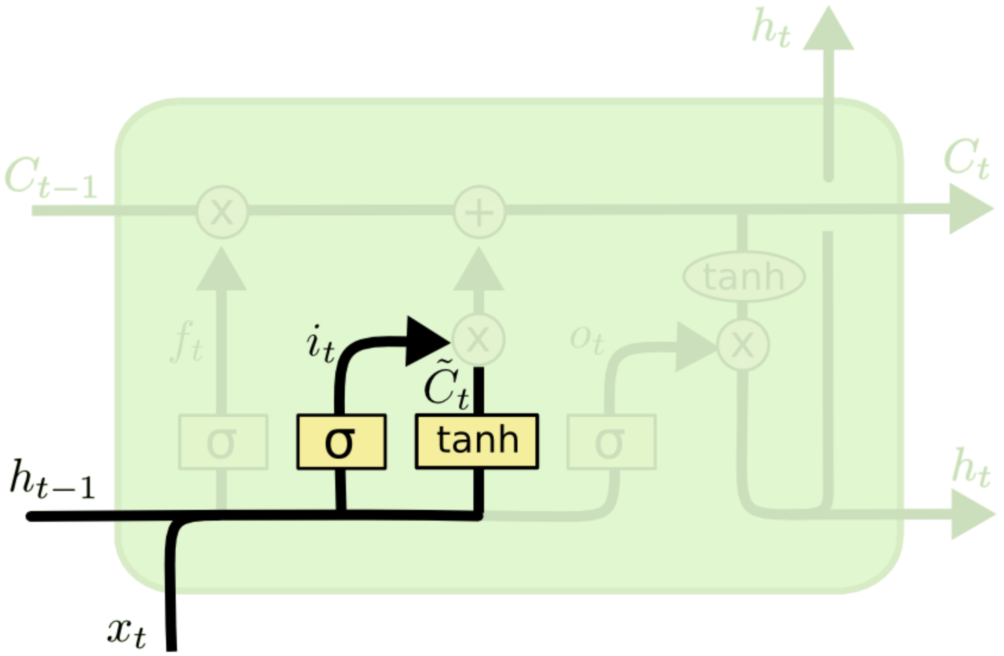


$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

LSTM



Фильтр входа:

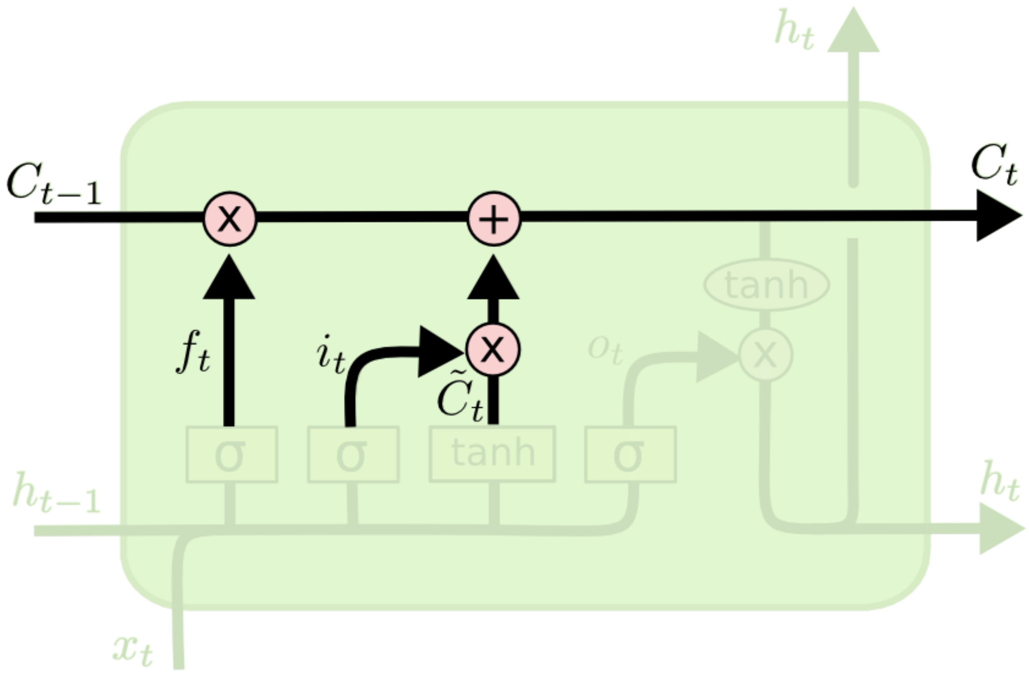


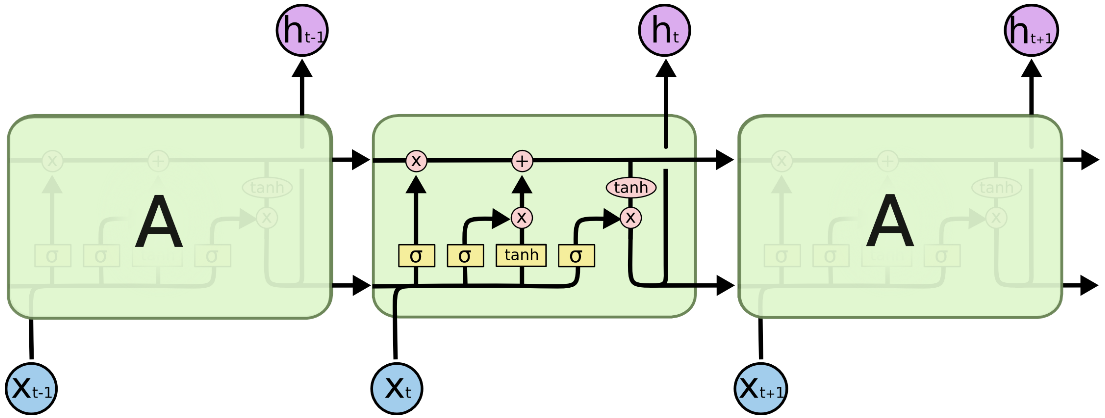
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$

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

LSTM:

Обновляем значения:

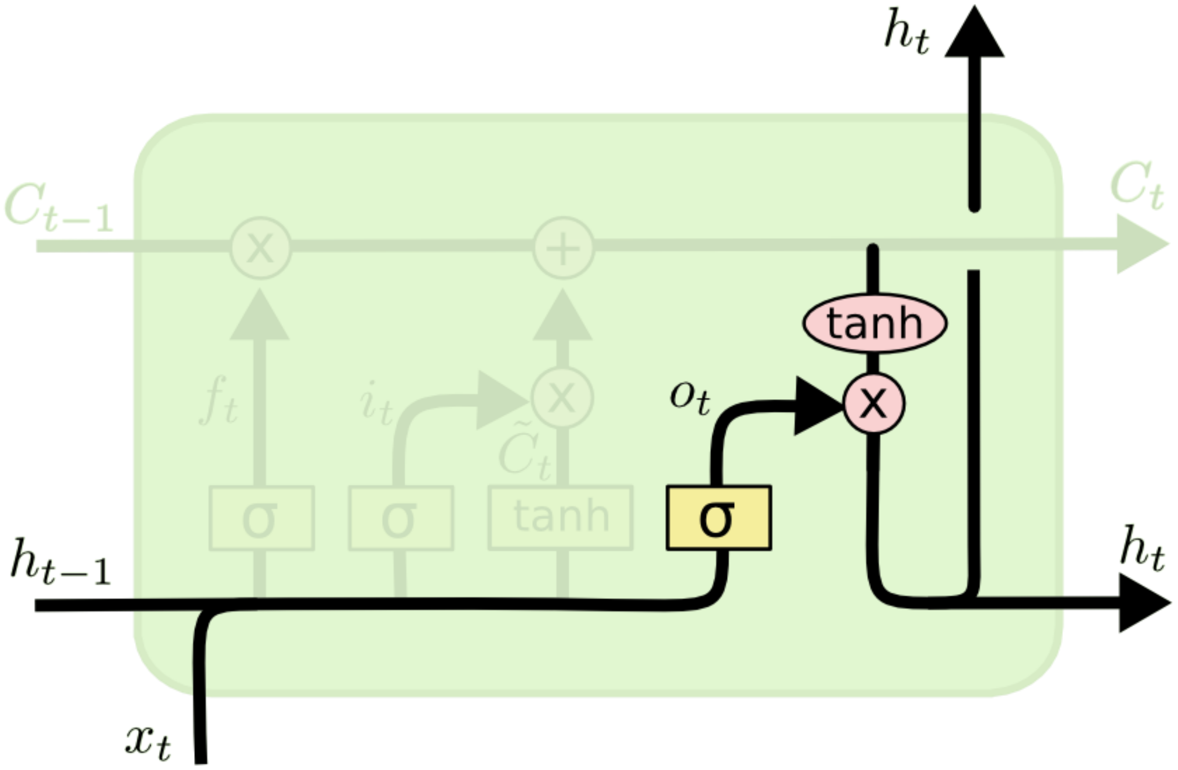


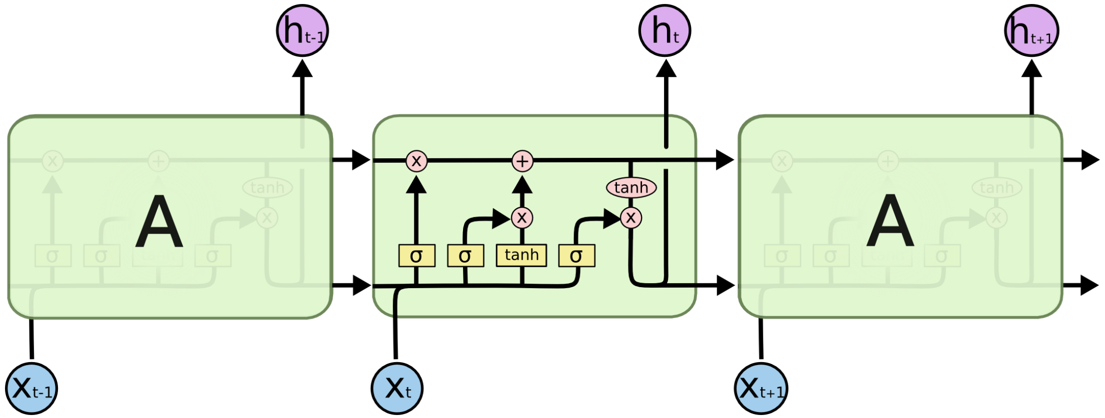


$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

LSTM

Фильтр выхода:



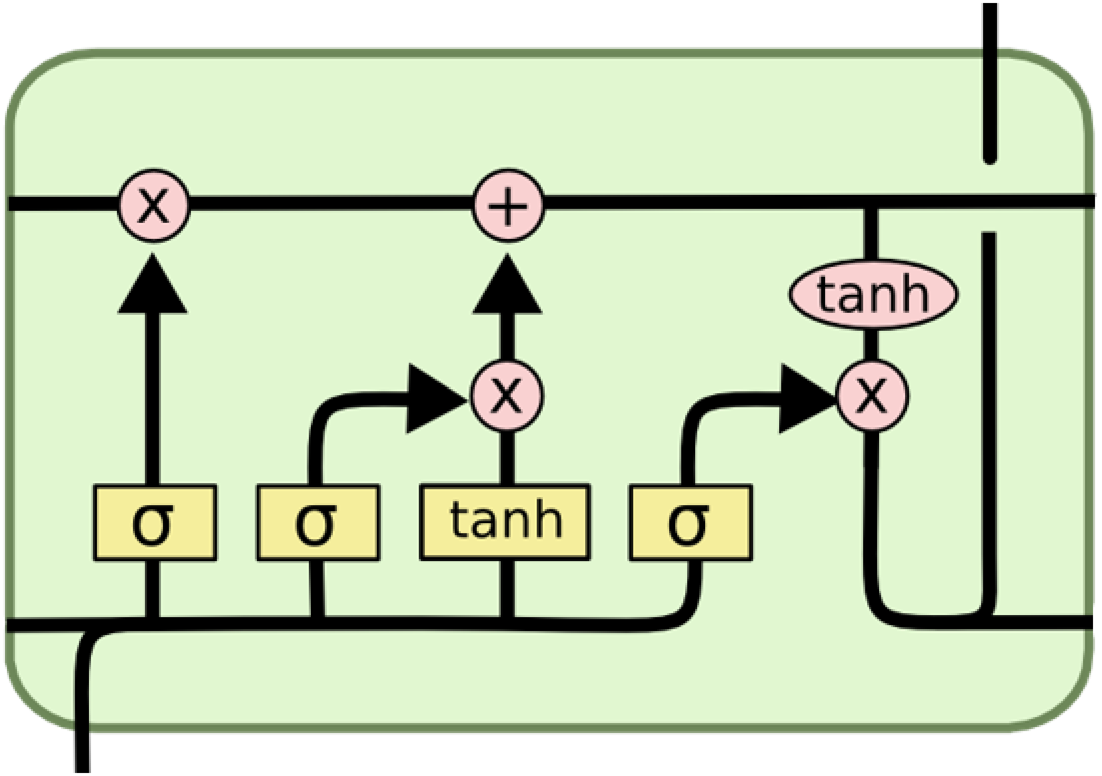


$$o_t = \sigma\left(W_o\ [h_{t-1}, x_t]\ +\ b_o\right)$$

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

LSTM

В итоге:



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

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$

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

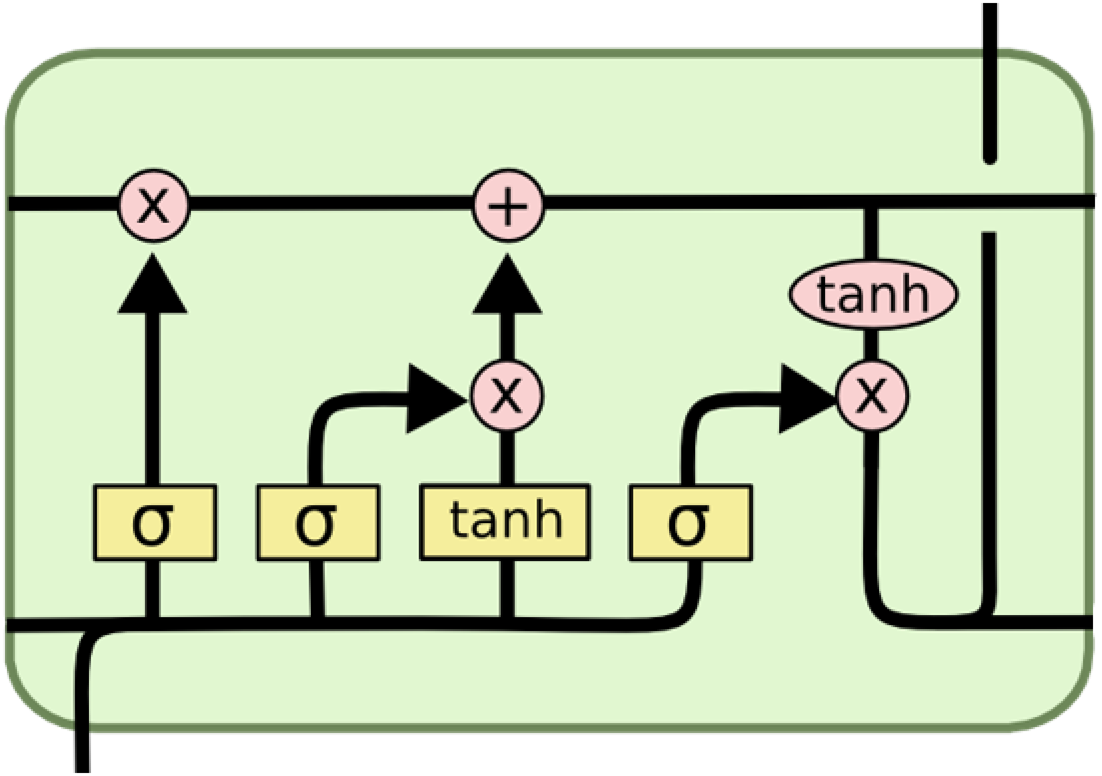
$$C_t = f_t * C_{t-1} + i_t * C_t$$

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

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

LSTM

В итоге:



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

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \;+\; b_i\right)$$

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

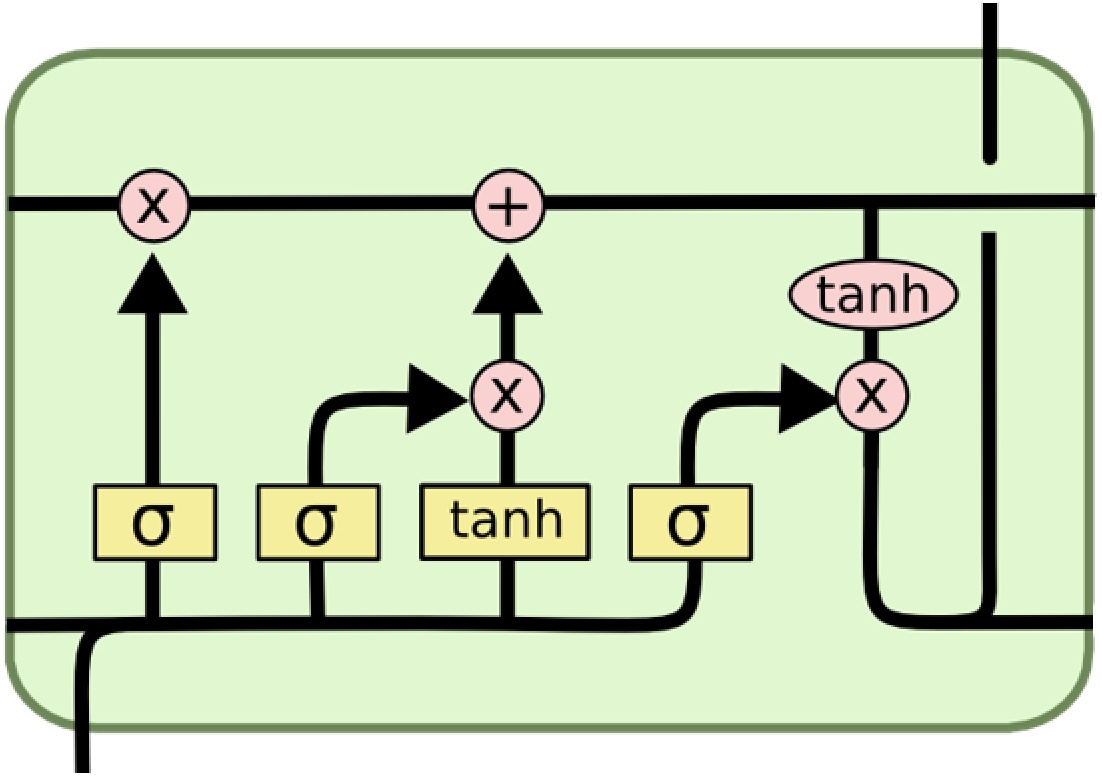
$$C_t = f_t * C_{t-1} + i_t * C_t$$

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

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

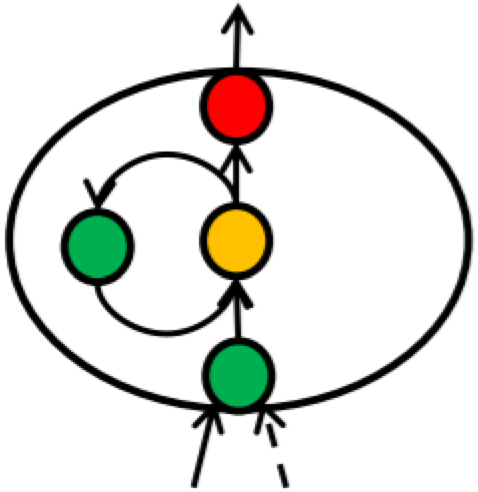
LSTM

В итоге:

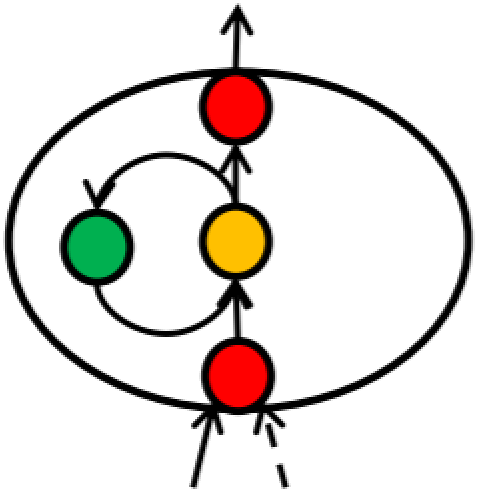


- Нашли путь, где градиенты долго не затухают
- $$\frac{\partial C_t}{\partial C_{t-1}} = f_t \Rightarrow$$ Большое начальное значение b_f

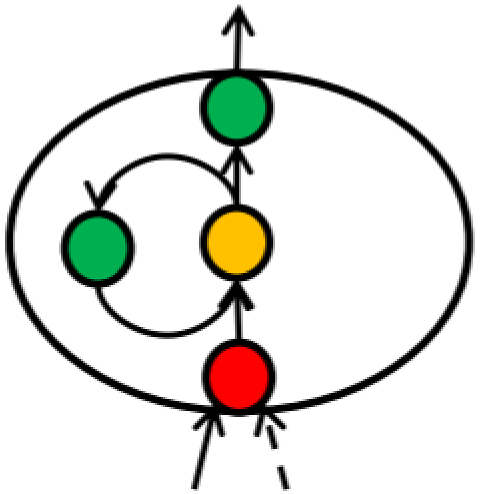
Как это работает



Запоминаем

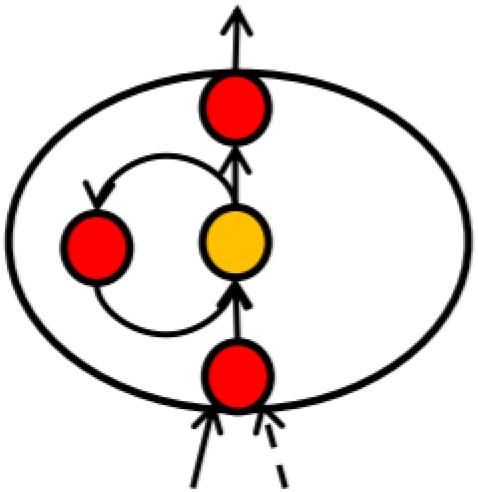


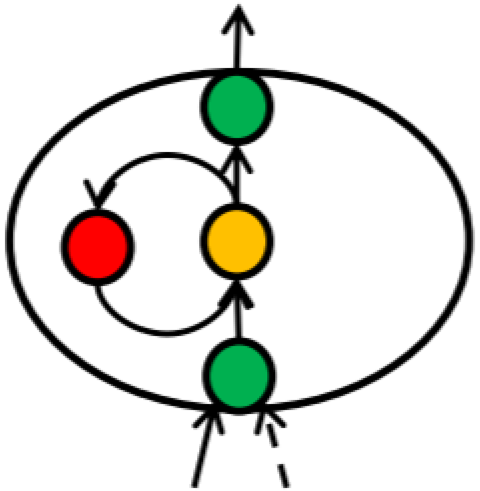
Обрабатываем



Отвечаем

Стираем



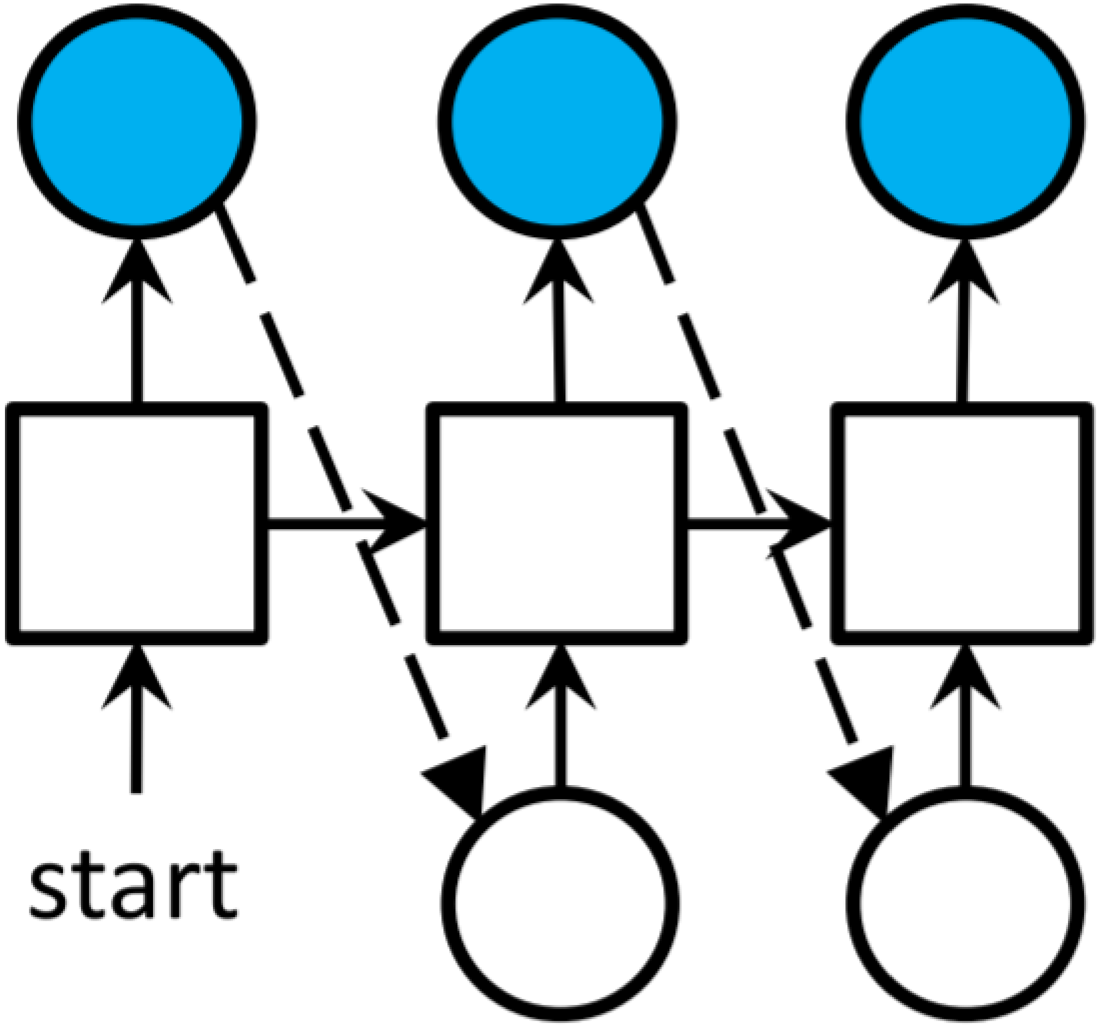


RNN

Генерация текста

Следующее слово:

Предыдущее слово:



Эксперименты Карпати. Алгебра

For $\bigoplus_{n=1,\dots,m}$ where $\mathcal{L}_{m_\bullet} = 0$, hence we can find a closed subset $\mathcal{H}$ in $\mathcal{H}$ and any sets $\mathcal{F}$ on X, U is a closed immersion of S, then $U \to T$ is a separated algebraic space.

Proof. Proof of (1). It also start we get

$$S = \mathrm{Spec}(R) = U \times_X U \times_X U$$

and the comparicoly in the fibre product covering we have to prove the lemma generated by $\coprod Z \times_U U \to V$. Consider the maps M along the set of points Sch_{fppf} and $U \to U$ is the fibre category of S in U in Section, **??** and the fact that any U affine, see Morphisms, Lemma **??**. Hence we obtain a scheme S and any open subset $W \subset U$ in $Sh(G)$ such that $\mathrm{Spec}(R') \to S$ is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that f_i is of finite presentation over S. We claim that $\mathcal{O}_{X,x}$ is a scheme where $x, x', s'' \in S'$ such that $\mathcal{O}_{X,x'} \to \mathcal{O}'_{X',x'}$ is separated. By Algebra, Lemma **??** we can define a map of complexes $\mathrm{GL}_{S'}(x'/S'')$ and we win. □

To prove study we see that $\mathcal{F}|_U$ is a covering of $\mathcal{X}'$, and $\mathcal{T}_i$ is an object of $\mathcal{F}_{X/S}$ for $i > 0$ and $\mathcal{F}_p$ exists and let $\mathcal{F}_i$ be a presheaf of $\mathcal{O}_X$-modules on $\mathcal{C}$ as a $\mathcal{F}$-module. In particular $\mathcal{F} = U/\mathcal{F}$ we have to show that

$$\widetilde{M}^\bullet = \mathcal{I}^\bullet \otimes_{\mathrm{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1}\mathcal{F})$$

is a unique morphism of algebraic stacks. Note that

$$\mathrm{Arrows} = (Sch/S)^{opp}_{fppf}, (Sch/S)_{fppf}$$

and

$$V = \Gamma(S, \mathcal{O}) \longmapsto (U, \mathrm{Spec}(A))$$

is an open subset of X. Thus U is affine. This is a continuous map of X is the inverse, the groupoid scheme S.

Proof. See discussion of sheaves of sets. □

The result for prove any open covering follows from the less of Example **??**. It may replace S by $X_{spaces,étale}$ which gives an open subspace of X and T equal to S_{Zar}, see Descent, Lemma **??**. Namely, by Lemma **??** we see that R is geometrically regular over S.

Lemma 0.1. *Assume (3) and (3) by the construction in the description.*

Suppose $X = \lim |X|$ (by the formal open covering X and a single map $\underline{Proj}_X(\mathcal{A}) =$ $\mathrm{Spec}(B)$ over U compatible with the complex

$$Set(\mathcal{A}) = \Gamma(X, \mathcal{O}_{X,\mathcal{O}_X}).$$

When in this case of to show that $\mathcal{Q} \to \mathcal{C}_{Z/X}$ is stable under the following result in the second conditions of (1), and (3). This finishes the proof. By Definition **??** *(without element is when the closed subschemes are catenary. If T is surjective we may assume that T is connected with residue fields of S. Moreover there exists a closed subspace $Z \subset X$ of X where U in X' is proper (some defining as a closed subset of the uniqueness it suffices to check the fact that the following theorem*

(1) *f is locally of finite type. Since $S = \mathrm{Spec}(R)$ and $Y = \mathrm{Spec}(R)$.*

Proof. This is form all sheaves of sheaves on X. But given a scheme U and a surjective étale morphism $U \to X$. Let $U \cap U = \coprod_{i=1,\dots,n} U_i$ be the scheme X over S at the schemes $X_i \to X$ and $U = \lim_i X_i$. □

The following lemma surjective restrocomposes of this implies that $\mathcal{F}_{x_0} = \mathcal{F}_{x_0} = \mathcal{F}_{\mathcal{X},\dots,0}$.

Lemma 0.2. *Let X be a locally Noetherian scheme over S, $E = \mathcal{F}_{X/S}$. Set $\mathcal{I} = \mathcal{J}_1 \subset \mathcal{I}'_n$. Since $\mathcal{I}^n \subset \mathcal{I}^n$ are nonzero over $i_0 \leq \mathfrak{p}$ is a subset of $\mathcal{J}_{n,0} \circ \overline{A}_2$ works.*

Lemma 0.3. *In Situation* **??**. *Hence we may assume $\mathfrak{q}' = 0$.*

Proof. We will use the property we see that $\mathfrak{p}$ is the mext functor (**??**). On the other hand, by Lemma **??** we see that

$$D(\mathcal{O}_{X'}) = \mathcal{O}_X(D)$$

where K is an F-algebra where δ_{n+1} is a scheme over S. □

Война и мир

100 итераций:

```
tyntd-iafhatawiaoihrdemot  lytdws  e ,tfti, astai f ogoh eoase rrranbyne 'nhthnee e
plia tklrgd t o idoe ns,smtt   h ne etie h,hregtrs nigtike,aoaenns lng
```

300 итераций:

```
"Tmont thithey" fomesscerliund
Keushey. Thom here
sheulke, anmerenith ol sivh I lalterthend Bleipile shuwy fil on aseterlome
coaniogennc Phe lism thond hon at. MeiDimorotion in ther thize."
```

500 итераций:

```
we counter. He stutn co des. His stanted out one ofler that concossions and was
to gearang reay Jotrets and with fre colt otf paitt thin wall. Which das stimn
```

Увеличение количества итераций

700 итераций:

```
Aftair fall unsuch that the hall for Prince Velzonski's that me of
her hearly, and behs to so arwage fiving were to it beloge, pavu say falling misfort
how, and Gogition is so overelical and ofter.
```

1200 итераций:

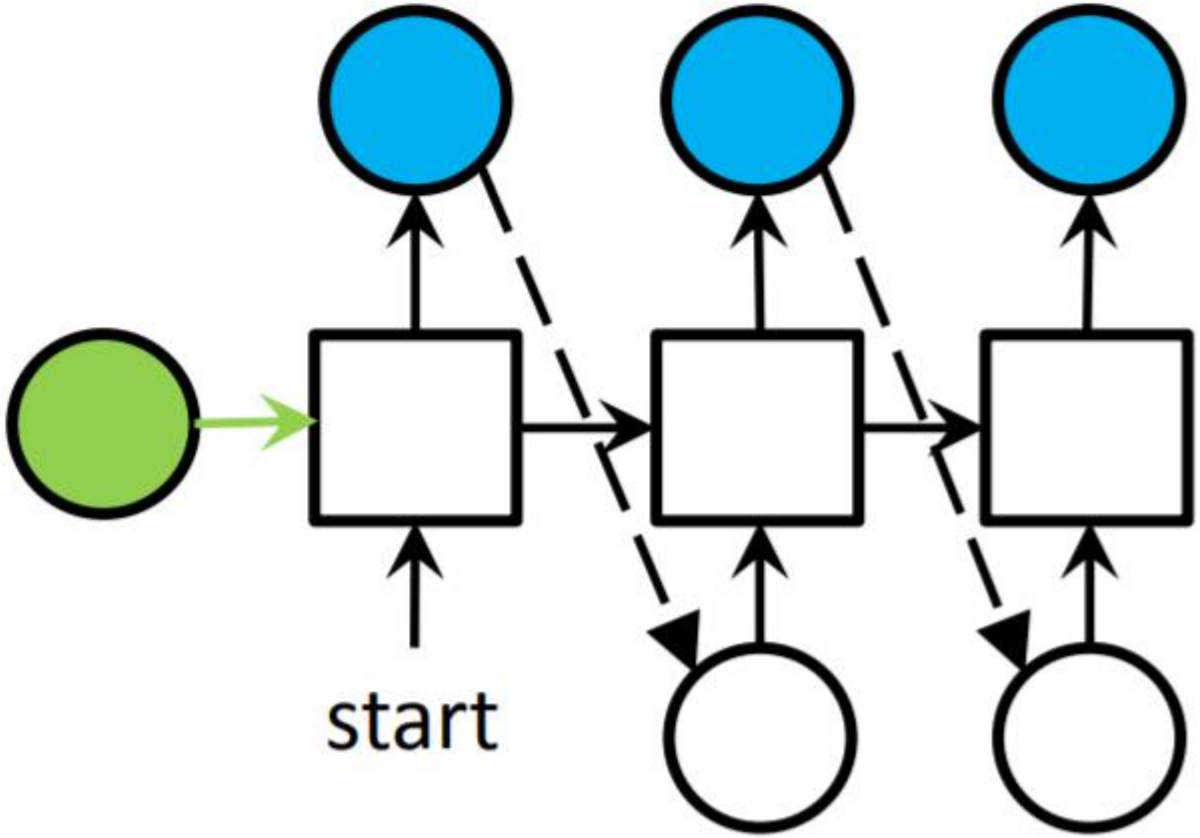
```
"Kite vouch!" he repeated by her
door. "But I would be done and quarts, feeling, then, son is people...."
```

2000 итераций:

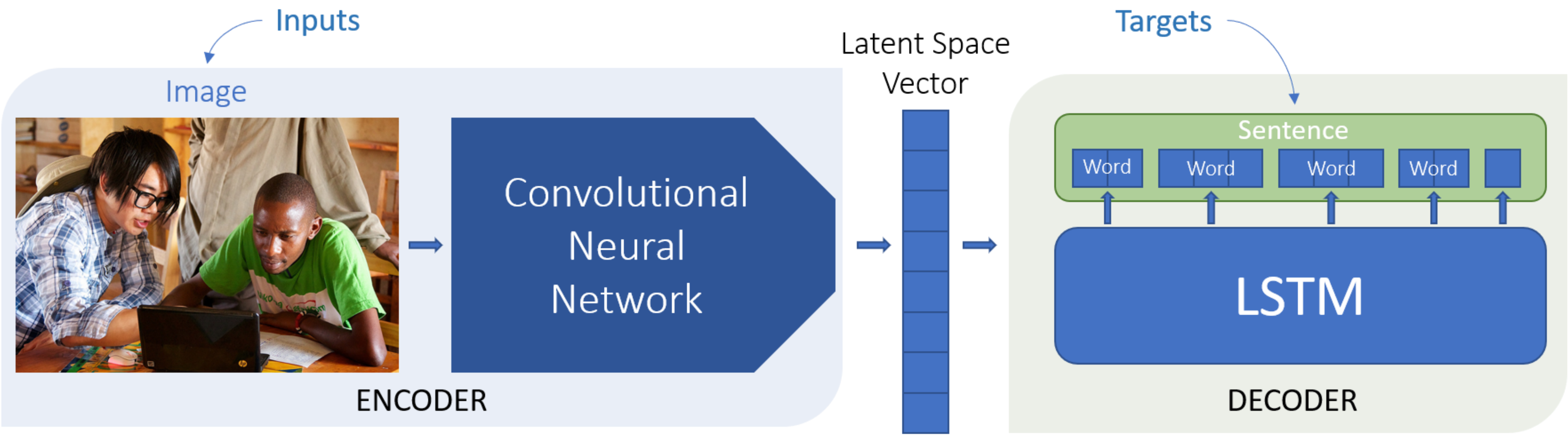
```
"Why do what that day," replied Natasha, and wishing to himself the fact the
princess, Princess Mary was easier, fed in had oftened him.
Pierre aking his soul came to the packs and drove up his father-in-law women.
```

Генерация подписей к картинкам

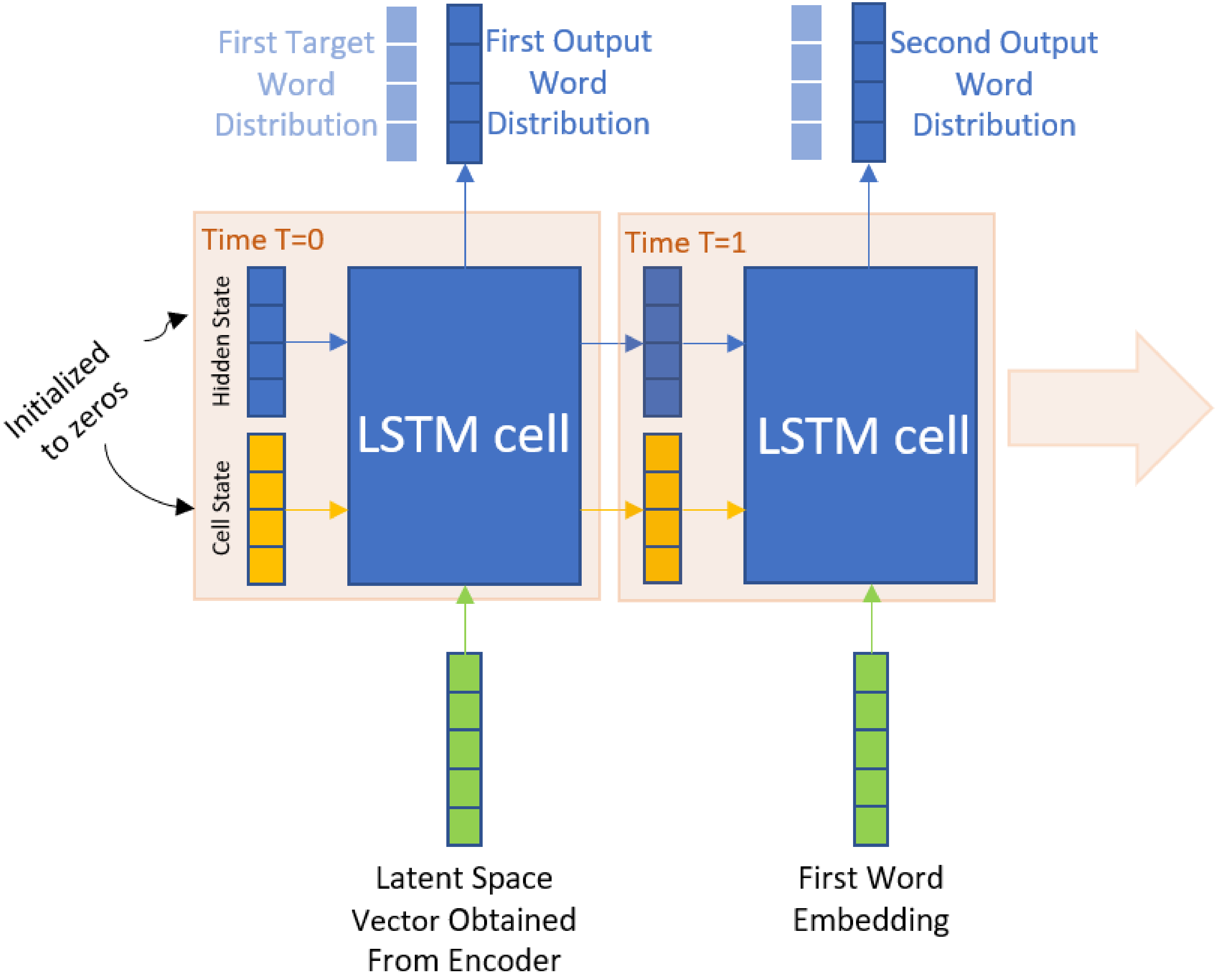
Как подать на вход RNN картинку?



Совмещение двух архитектур



Взглянем на LSTM

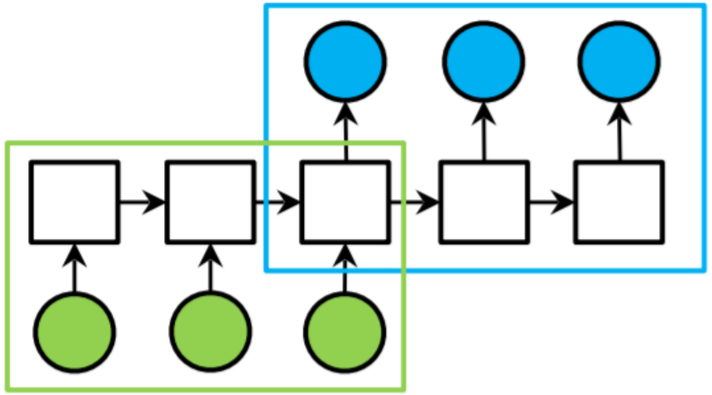


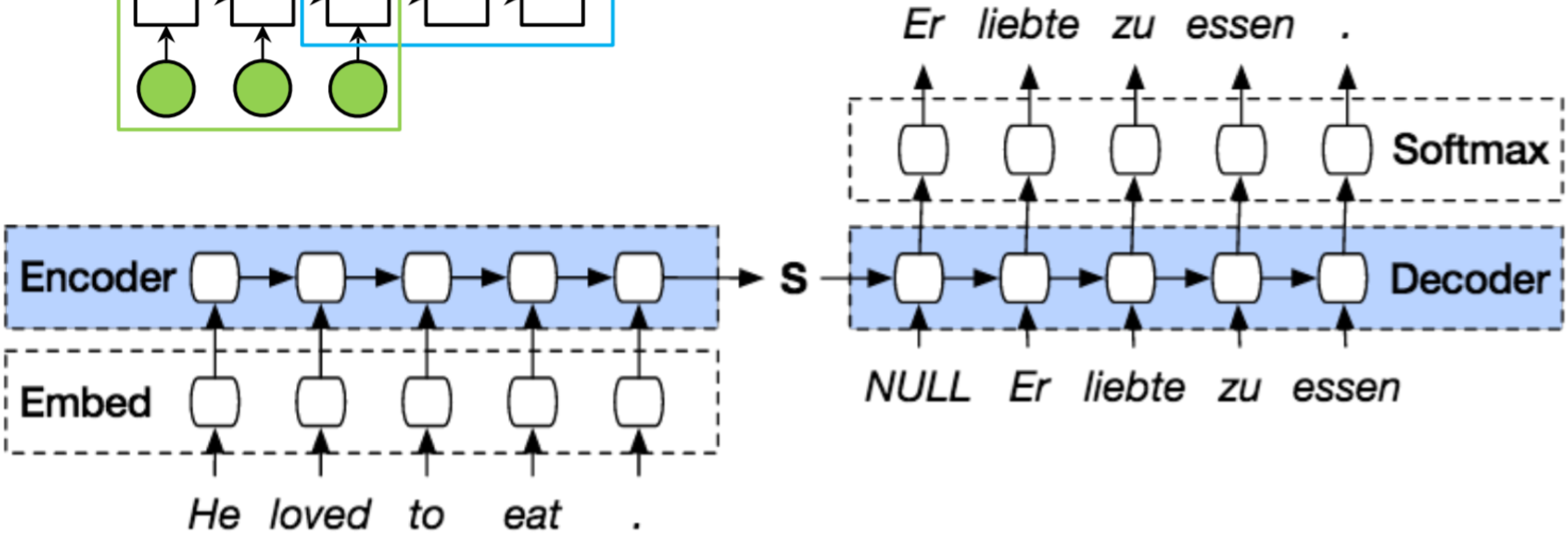
Результаты

people on skis on a snowy road with mountains in the background .

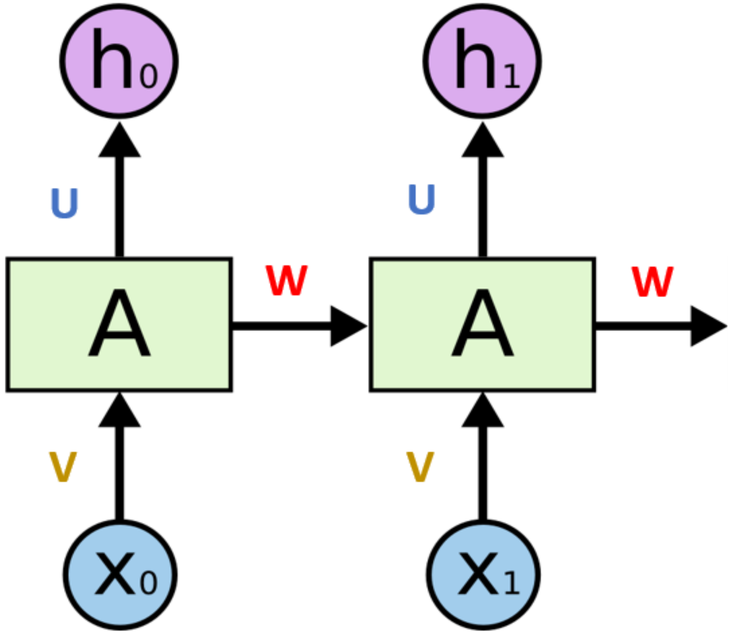
zebra grazing on lush green grass in the woods

Машинный перевод

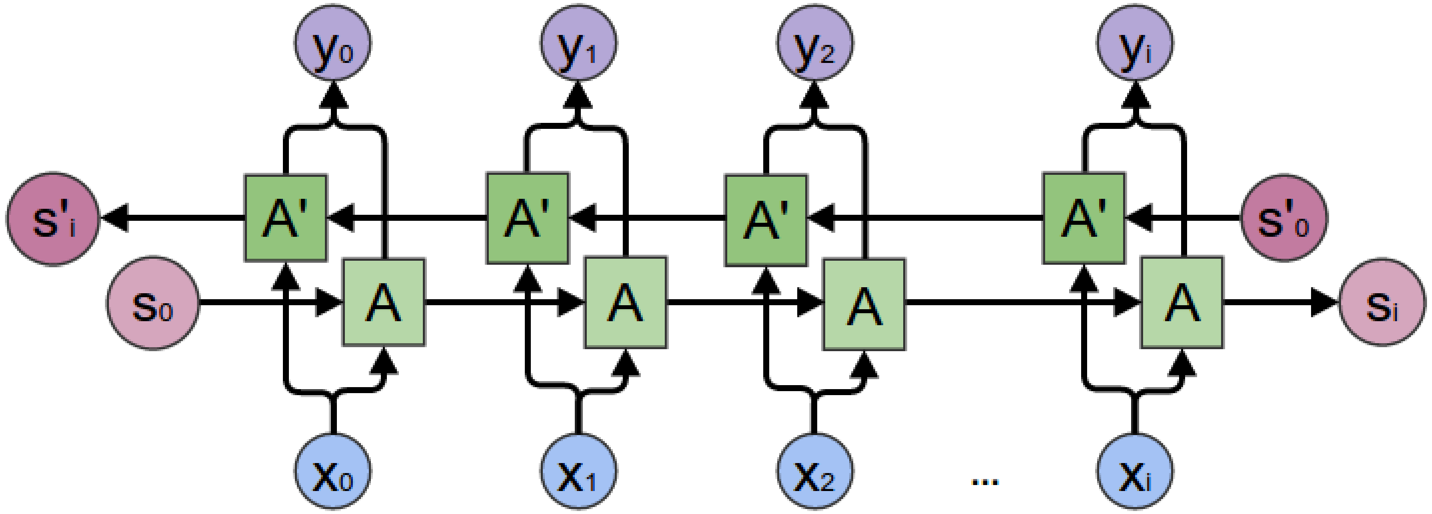




Двунаправленная RNN



Хорошо работает для задач, связанных с языком

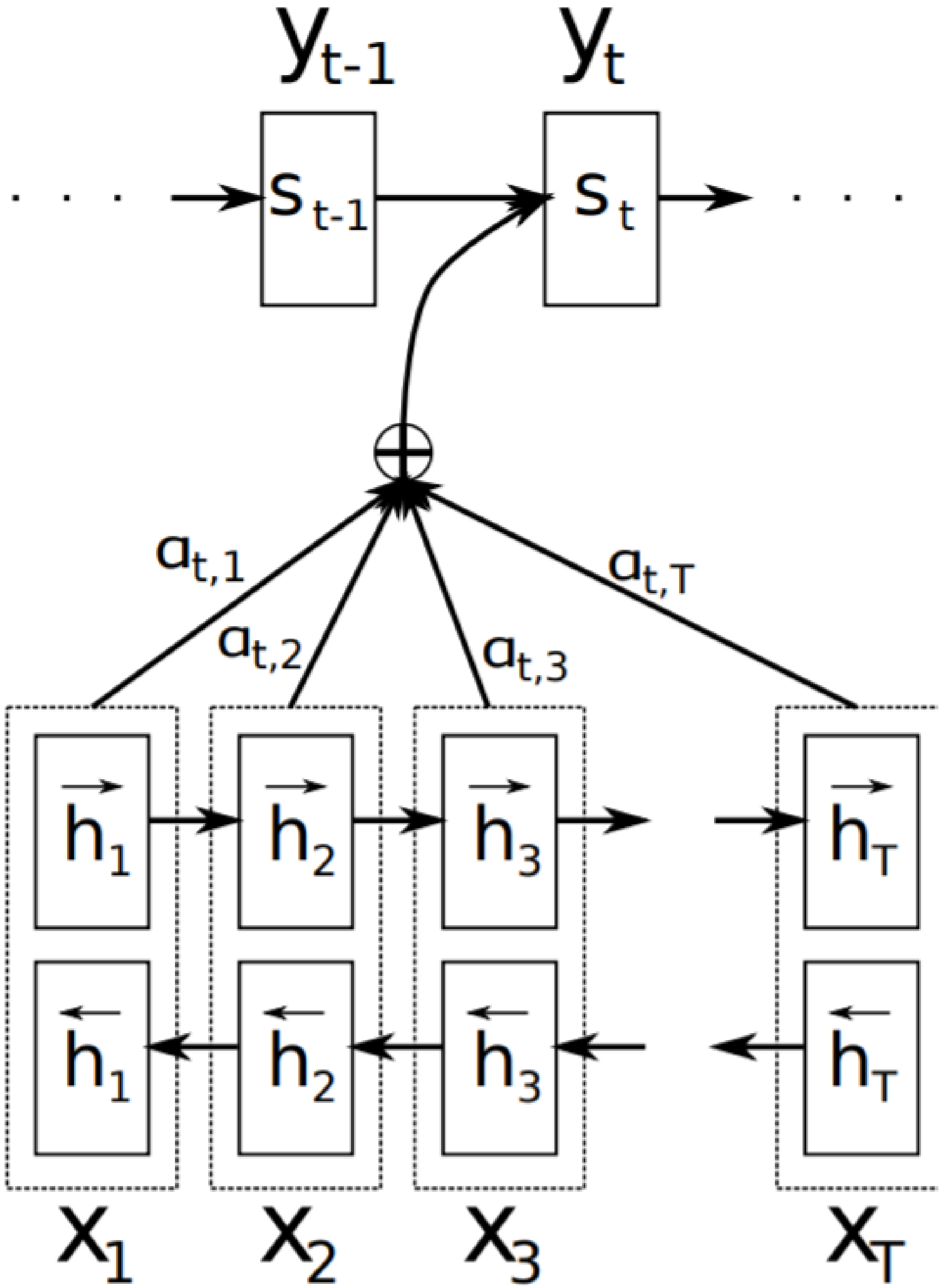


Двунаправленная RNN для машинного перевода

- Позволяет не генерировать вектор фиксированной длины для каждого предложения
- Лучше работает с длинными предложениями

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j \qquad \alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})},$$

$$s_i = f(s_{i-1}, y_{i-1}, c_i) \quad e_{ij} = a(s_{i-1}, h_j)$$



Выводы

- RNN действительно круты в анализе последовательностей
- Сами по себе обучаются сложно, градиенты взрываются или затухают, нужны модификации
- Они есть: LSTM или GRU, с их помощью запоминаем долгосрочные зависимости

Вопросы

- В чем идея LSTM? В чем отличие от обычной RNN?
- Как можно применять RNN для генерации подписей к изображениям?
- Зачем нужна двунаправленная RNN?

Источники

- https://arxiv.org/pdf/1409.0473.pdf
- https://arxiv.org/pdf/1409.3215.pdf
- https://compsciclub.ru/media/courses/2016-summer/spb-deep-learning/slides/deep_learning_lecture_230716.pdf
- http://karpathy.github.io/2015/05/21/rnn-effectiveness/
- http://colah.github.io/posts/2015-08-Understanding-LSTMs/