

Нейронные сети. Рекуррентные нейронные сети

Алексей Колесов

Белорусский государственный университет

2 декабря 2018 г.

Содержание

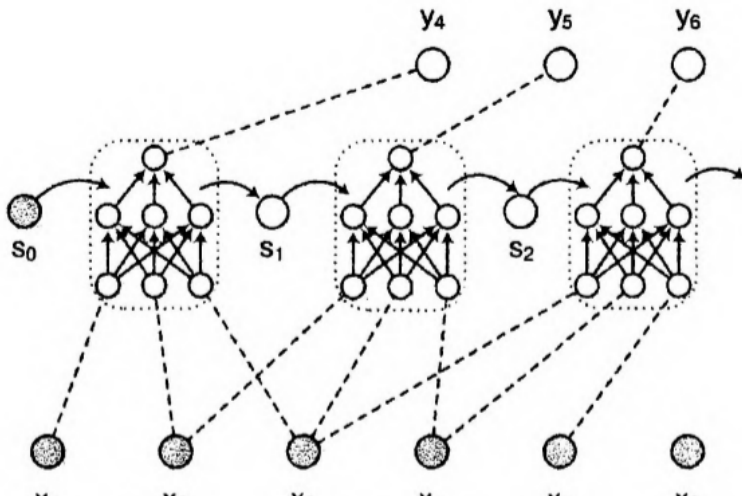
1 Рекуррентные нейронные сети

- Обработка последовательностей
- Backprop through time
- Long short-term memory
- Gated recurrent unit
- Замечания

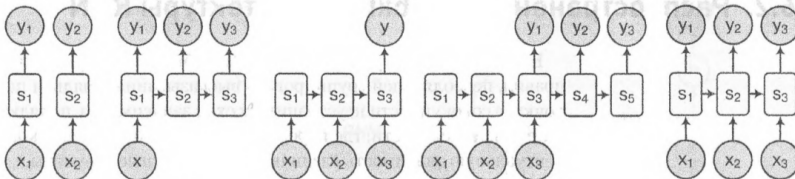
Обработка последовательностей

- до этого на входе одна точка x , на выходе — одна точка y
- что делать, если вход $x = (x_1, \dots, x_l)$, а $y = (y_1, \dots, y_l)$ где l — неизвестно?
- вариант: y_i зависит только от (x_{i-k}, \dots, x_i) , где i — константа

RNN



Задачи на последовательности

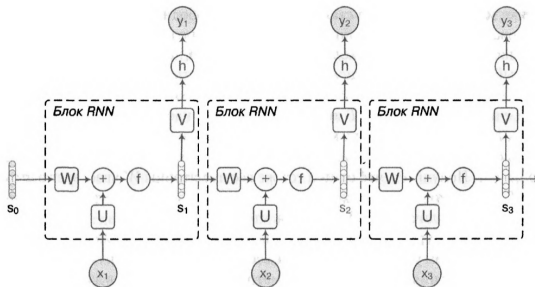


Содержание

1 Рекуррентные нейронные сети

- Обработка последовательностей
- **Backprop through time**
- Long short-term memory
- Gated recurrent unit
- Замечания

Simple RNN



$$a_t = b + Ws_{t-1} + Ux_t$$

$$o_t = c + Vs_t$$

$$s_t = f(a_t)$$

$$y_t = h(o_t)$$

Где глубина?

Можно сделать глубокими следующие преобразования:

- вход \Rightarrow состояние
- состояние \Rightarrow выход
- состояние \Rightarrow следующее состояние
- выход одного слоя подавать другому рекуррентному слою

Проблемы с BPTT

- взрыв градиентов — gradient clipping
- затухание градиентов — другие архитектуры

Между слоями лучше добавлять residual connections

Содержание

1 Рекуррентные нейронные сети

- Обработка последовательностей
- Backprop through time
- **Long short-term memory**
- Gated recurrent unit
- Замечания

LSTM

$$c'_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_{c'})$$

candidate cell state

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$

input gate

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

forget gate

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o)$$

output gate

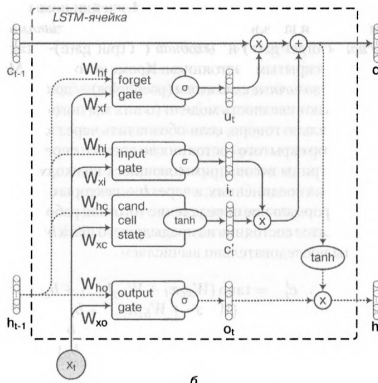
$$c_t = f_t \odot c_{t-1} + i_t \odot c'_t$$

cell state

$$h_t = o_t \odot \tanh(c_t)$$

block output

LSTM



Содержание

1 Рекуррентные нейронные сети

- Обработка последовательностей
- Backprop through time
- Long short-term memory
- **Gated recurrent unit**
- Замечания

GRU

$$u_t = \sigma(W_{xu}x_t + W_{hu}h_{t-1} + b_u)$$

update gate

$$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r)$$

reset gate

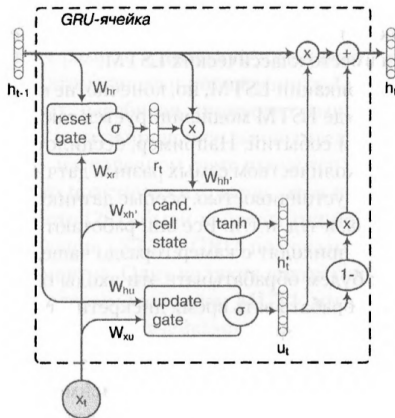
$$h'_t = \tanh(W_{xh'}x_t + W_{hh'}(r_t \odot h_{t-1}))$$

candidate state/output

$$h_t = (1 - u_t) \odot h'_t + u_t \odot h_{t-1}$$

state/output

GRU



Содержание

1 Рекуррентные нейронные сети

- Обработка последовательностей
- Backprop through time
- Long short-term memory
- Gated recurrent unit
- Замечания

Замечания

- учить — сложно; используйте gradient clipping, правильную инициализацию, residual connections
- используйте LSTM/GRU
- используйте bidirectional rnns
- используйте CNN

Итоги

- рассмотрели устройство рекуррентных нейронных сетей

Литература

- С. Николенко, А. Кадури́н, Е. Архангельская — Глубокое обучение (глава 6)