

# Recurrent Neural Network

Пудяков Ярослав

Национальный Исследовательский Университет  
Высшая Школа Экономики

*yaarpudyakov@edu.hse.ru*

17 ноября 2018 г.

# План презентации

## 1 Введение

- Данные, которые хотим обрабатывать (sequential data)
- Какие задачи хотим решать? Постановка задач.

## 2 Структура RNN сети.

- Общая схема RNN. Блочная структура.
- Разбор структуры блока.
- Математическое представление RNN

## 3 Работа с математической моделью

- Параметры модели. Что обучаем?
- Функция ошибки и ее минимизация
- Процесс обучения. Обратное распространение ошибки

## 4 Усовершенствование модели

- RNN с несколькими слоями. Как стекаем слои.
- Почему такой подход работает?
- Bidirectional RNN

## 5 Работа RNN на примерах

# Sequential Data

## • Стоимость акций



## • Текст

Univers 95 Roman

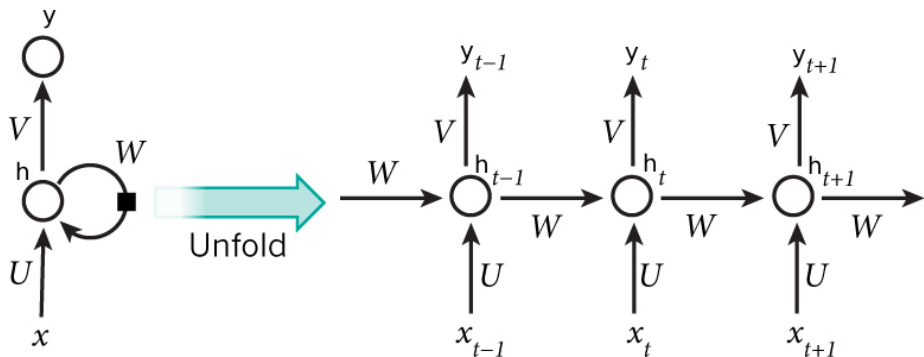
Lorem ipsum dolor sit amet, magna consetetur  
sadipscing elitr, sed diam nonumy eirmod  
tempor invidunt ut labore et dolore in magna  
aliquyam erat, sed diam voluptua. At vero  
eos et accusam et justo duo dolores et ea rebus.  
Stet clita kasd gubergren, no sea takimata  
sanctus est Lorem ipsum dolor sit amet. Lorem  
ipsum dolor sit amet, consetetur sadipscing  
elitr, sed diam nonumy eirmod tempor invidunt  
ut labore et dolore magna aliquyam erat, sed  
diam voluptua. At vero eos et accusam et justo  
duo dolores et ea rebum. Stet clita kasd  
gubergren, no sea takimata sanctus est Lorem  
ipsum dolor sit amet. Lorem ipsum dolor sit  
amet, consetetur sadipscing non elitr, sed diam  
nonumy eirmod tempor invidunt ut labore et  
dolore magna aliquyam erat, sed diam sanctus  
voluptua.

## • Звук



- Прогноз стоимости акций на следующий день.
- Классификация текста по эмоциональной окраске (положительный/отрицательный).
- Машинный перевод на другой язык.
- Распознавание речи.
- Написание рассказа.
- ...

# Блочная структура RNN. Общая схема.

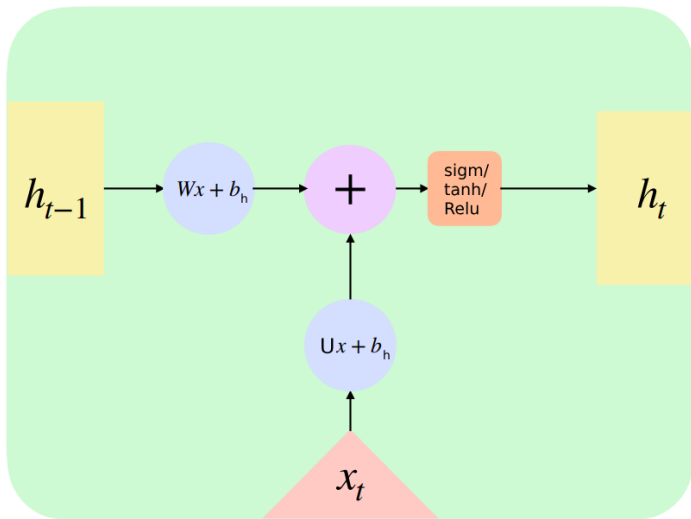


$$h_t = f_h(Wh_{t-1} + Ux_t + b_h)$$

$$y_t = f_y(Vh_t + b_y)$$

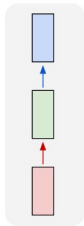
[<http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/>]

# Разбор структуры блока

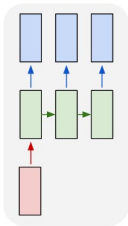


# Вариации модели

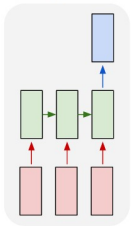
one to one



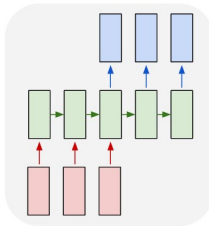
one to many



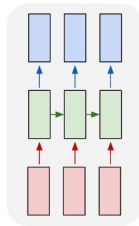
many to one



many to many



many to many



[<https://www.microsoft.com/en-us/cognitive-toolkit/blog/2016/11/sequence-to-sequence-deep-recurrent-neural-networks-in-cntk-part-1/>]

$$h_t = f_h(Wh_{t-1} + Ux_t + b_h)$$

$$y_t = f_y(Vh_t + b_y)$$

$f_h$  – *sigm/tanh/Relu/LeakRelu...*

$f_y$  – например *softmax*

$W$  - матрица весов скрытого слоя.

$U$  - матрица весов входного слоя.

$V$  - матрица весов выходного слоя

$b_h, b_y$  - векторные сдвиги для скрытого и выходного слоев



# Функция ошибки

$\hat{y}_i$  - предсказание

$y_i$  - истинное значение

$L_i(y_i, \hat{y}_i)$  - функция ошибки для выхода  $y_i$

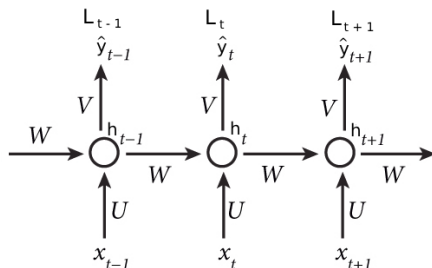
Примеры L:

- $L(y, \hat{y}) = -\frac{1}{N} \sum_{n \in N} (y_n) \log(\hat{y}_n)$
- $L(y, \hat{y}) = \frac{1}{N} \sum_{n \in N} (\hat{y}_n - y_n)^2$
- ...

Необходимо  $L(y, \hat{y}) \rightarrow \min$

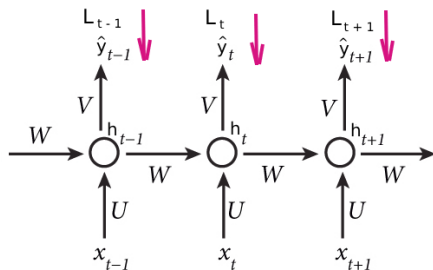
Обучаем градиентным спуском.

# Функция ошибки



Необходимо посчитать:  $\frac{\partial L}{\partial U}, \frac{\partial L}{\partial V}, \frac{\partial L}{\partial W}, \frac{\partial L}{\partial b_h}, \frac{\partial L}{\partial b_y}$

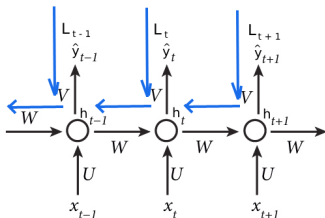
# Функция ошибки



$$\frac{\partial L}{\partial V} = \sum_{i=0}^T \frac{\partial L_i}{\partial V} = \sum_{i=0}^T \frac{\partial L_i}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial V}$$

$$\hat{y}_i = f_y(Vh_t + b_y)$$

# Функция ошибки

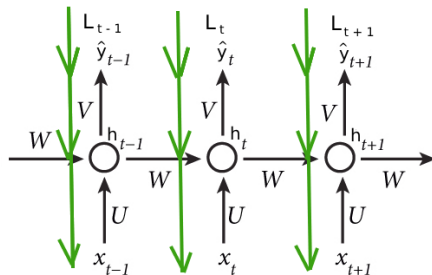


$$\frac{\partial L}{\partial W} = \sum_{i=0}^T \frac{\partial L_i}{\partial W} = \sum_{i=0}^T \frac{\partial L_i}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial W} = \sum_{i=0}^T \frac{\partial L_i}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial h_t} \frac{\partial h_t}{\partial W}$$

$$h_t = f_h(Ux_t + Wh_{t-1} + b_h)$$

$$\frac{\partial L}{\partial W} = \sum_{i=0}^T \frac{\partial L_i}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial h_t} \left( \frac{\partial h_t}{\partial W} + \frac{\partial h_t}{\partial h_{t-1}} \frac{\partial h_{t-1}}{\partial W} + \dots \right)$$

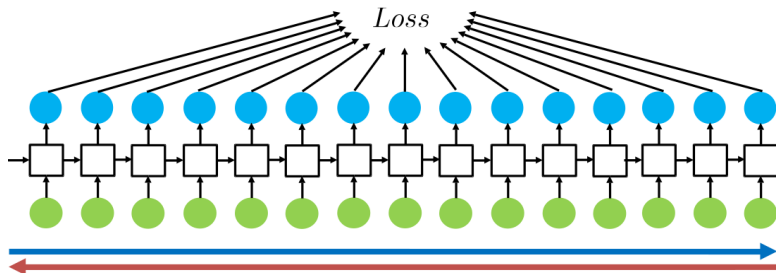
# Функция ошибки



$$\frac{\partial L}{\partial U} = \sum_{i=0}^T \frac{\partial L_i}{\partial U} = \sum_{i=0}^T \frac{\partial L_i}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial U} = \sum_{i=0}^T \frac{\partial L_i}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial h_t} \frac{\partial h_t}{\partial U}$$

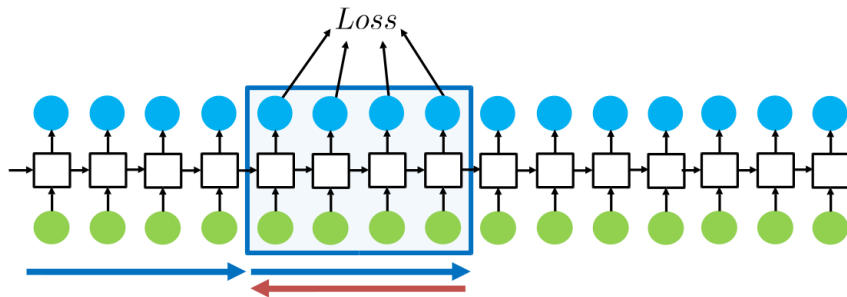
$$\hat{y}_i = f_y(Vh_t + b_y)$$

# Truncated BPTT



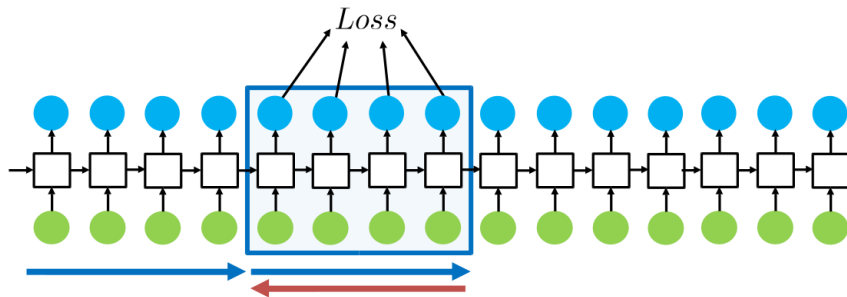
[DeepBayes lecture: Recurrent neural networks, Ekaterina Lobacheva, 2017]

# Truncated BPTT



[DeepBayes lecture: Recurrent neural networks, Ekaterina Lobacheva, 2017]

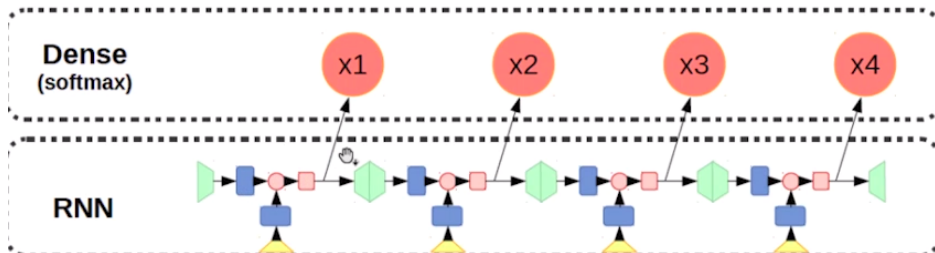
# Truncated BPTT



[DeepBayes lecture: Recurrent neural networks, Ekaterina Lobacheva, 2017]

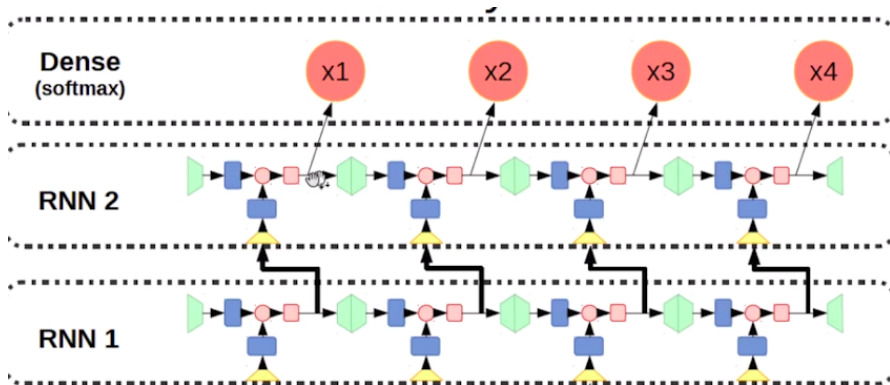


# Усовершенствование однослойной RNN



[ШАД: Рекуррентные нейронные сети, О. Васильев, 2017]

# Добавление RNN слоя

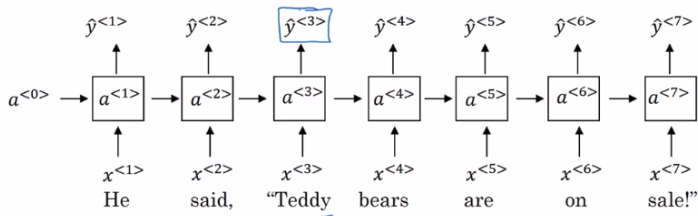


[ШАД: Рекуррентные нейронные сети, О. Васильев, 2017]

## Getting information from the future

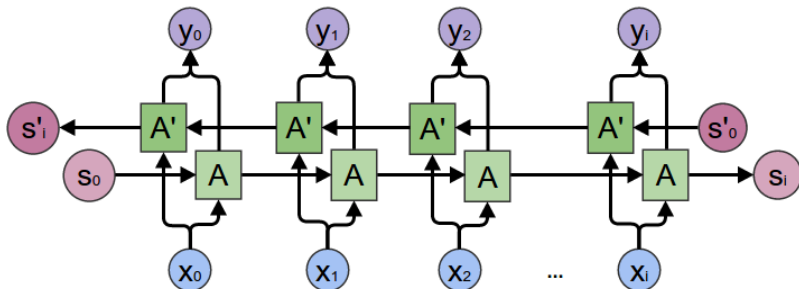
He said, "Teddy bears are on sale!"

He said, "Teddy Roosevelt was a great President!"



[Coursera: Sequence Model (Bidirectional RNN), Andrew Ng]

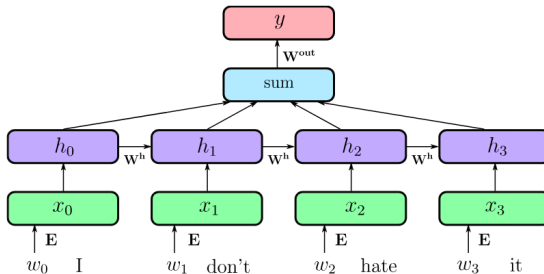
# Bidirectional RNN. Общая схема



[<https://towardsdatascience.com/introduction-to-sequence-models-rnn-bidirectional-rnn-lstm-gru-73927ec9df15>]

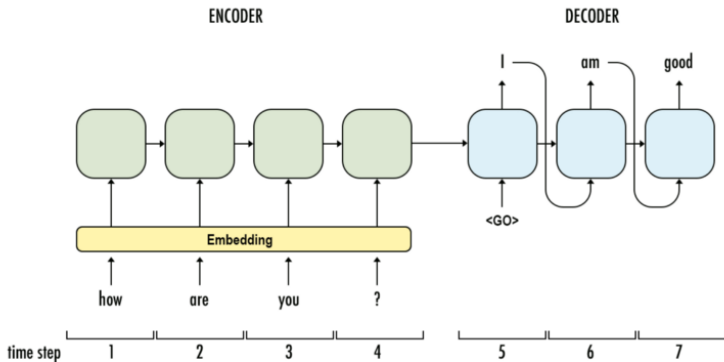
$$y_t = f_y(V[a_t, a'_t] + b_y)$$

# Работа RNN на примерах. Sentinment Analysis



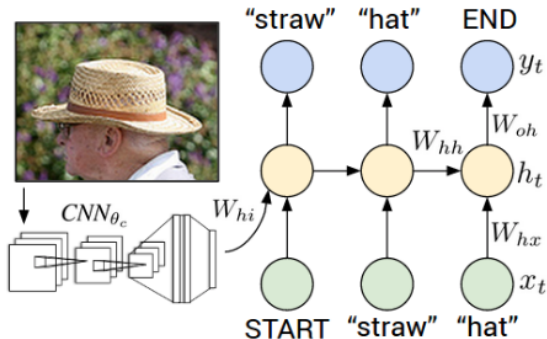
[Natural Language Processing with Deep Learning: Charles Ollion, Olivier Grisel, 2017]

# Работа RNN на примерах. Dialogue System



[<https://medium.com/syncedreview/a-brief-overview-of-attention-mechanism-13c578ba9129>]

# Работа RNN на примерах. Image Captioning



[<https://towardsdatascience.com/image-captioning-in-deep-learning-9cd23fb4d8d2>]

- RNN применяются для работы с данными, представленными в виде последовательностей. (генерация текстов, классификация музыки, машинный перевод и тд.)
- RNN представляет из себя блочную схему, в которой фигурируют скрытые состояния, входные и выходные данные. С этими данными происходят преобразования, задающиеся матрицами и векторами.
- Процесс обучения сводится к нахождению оптимальных весов матриц  $U, W, V, b_h, b_x$ , которые минимизируют функцию ошибки. Для этого используется градиентные спуск.
- Ошибку в определенный момент можно считать не для всех выходов, а для последних  $K$  - это называется Truncated BPTT.
- Для улучшения качества работы можно применять многослойные RNN.
- Bidirectional RNN применяются тогда, когда нам необходимо знать информацию о следующих входах ( $x_i$ ), а не только о предыдущих.



- 1 DeepBayes lecture: Recurrent neural networks, Ekaterina Lobacheva, 2017
- 2 ШАД: Рекуррентные нейронные сети, О. Васильев, 2017
- 3 Coursera: Sequence Model (Bidirectional RNN), Andrew Ng
- 4 Natural Language Processing with Deep Learning: Charles Ollion, Olivier Grisel, 2017