

академия
больших
данных



mail.ru
group



Хорошо ли видно и слышно?

Поднимите руку или напишите о проблеме

Проверить запись



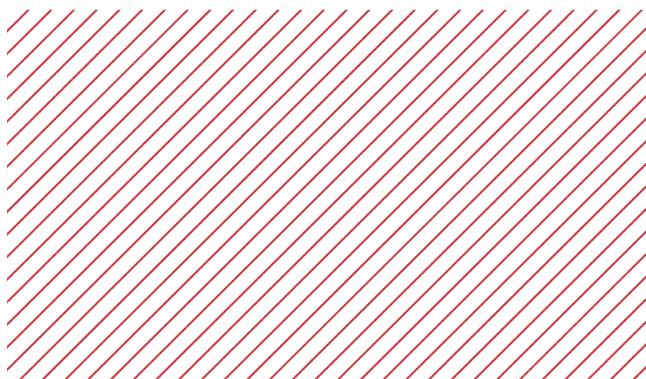
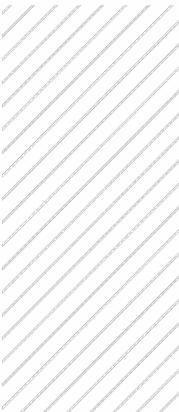
академия
больших
данных



Распознавание текста

Иван Карпухин

Ведущий программист-исследователь в
команде машинного зрения



Преподаватель



Карпухин Иван

Профессионально занимаюсь машинным обучением
более 5 лет

Проекты (Яндекс, Mail.ru):

- Голосовая биометрия для Алисы
- Распознавание лиц и текстов
- Визуальный аватар



karpuhini@yandex.ru

Формат лекции



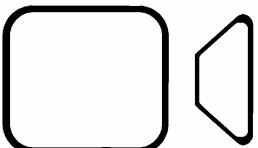
Активно участвуем

Задаем вопросы в чат или голосом 



Вопросы вижу в чате, могу ответить не сразу

Off-topic обсуждаем в Discord



+1 к карме за включенную камеру)

Формат лекции



План лекции





Цели

1

Получить теорию для семинара

2

Научиться вычислять CTC loss

3

Вспомнить архитектуры сетей для OCR

Вы их уже знаете!)

Пайплин OCR



Задание

Приведите примеры текстов из реальной жизни

 Пишите в чат или называйте голосом

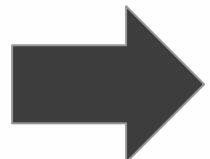
 1 минута



This message is secret 1234
~~SECRET~~ ~~SECRET~~ ~~SECRET~~ ~~SECRET~~ ~~SECRET~~ ~~SECRET~~ ~~SECRET~~

Схема распознавания текста

Считаем, что каждое слово (или буква) хорошо описывается прямоугольником



Детектирование

Распознавание
Исправление опечаток

Liberals
Please continue on 1-40
until you have left our
GREAT STATE OF
TEXAS
BURKETT

Разметка для обучения



Положение фрагментов
(BBox, mask)

Liberals

“Liberals”

our

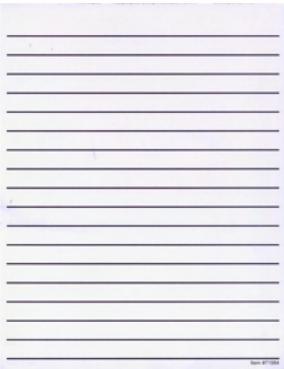
“our”

I-40

“I-40”

Текст каждого фрагмента

Синтетические данные



Фон



Times New Roman
Times New Roman Italic
Times New Roman Bold
Times New Roman Bold Italic

Шрифт



Текст



Аугментации

Вопросы

Детектирование

Детектирование текста

Какие архитектуры для детектирования вы помните?

💬 Пишите в чат



Детектирование текста

Первый путь: детектировать bounding boxes

Примеры: R-CNN, YOLO



Детектирование текста

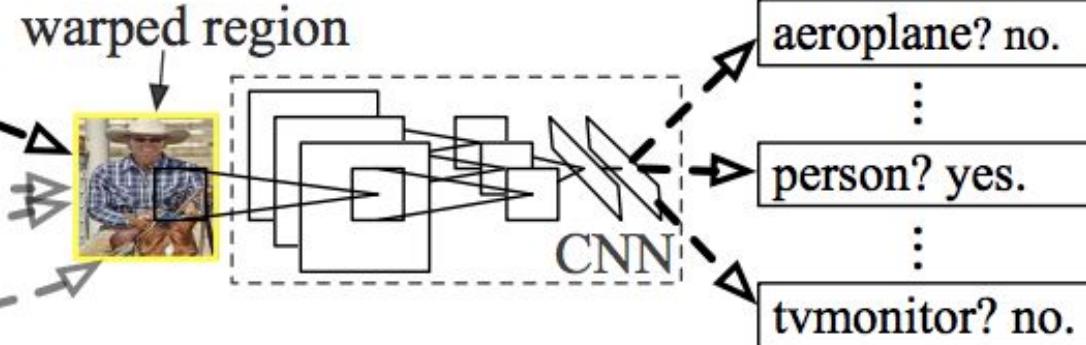
R-CNN: *Regions with CNN features*



1. Input
image



2. Extract region
proposals (~2k)



3. Compute
CNN features

4. Classify
regions

Проблема с BBox

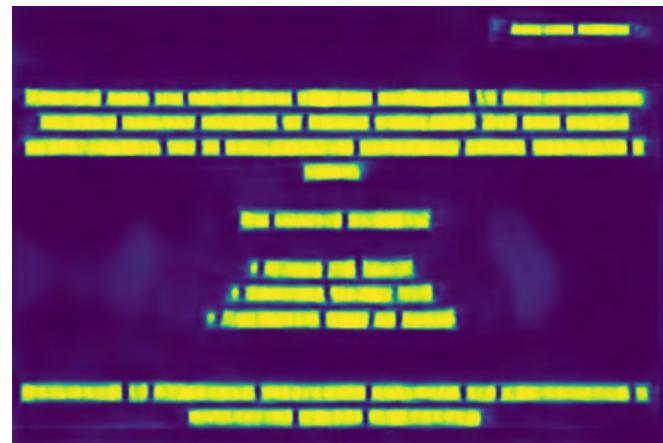
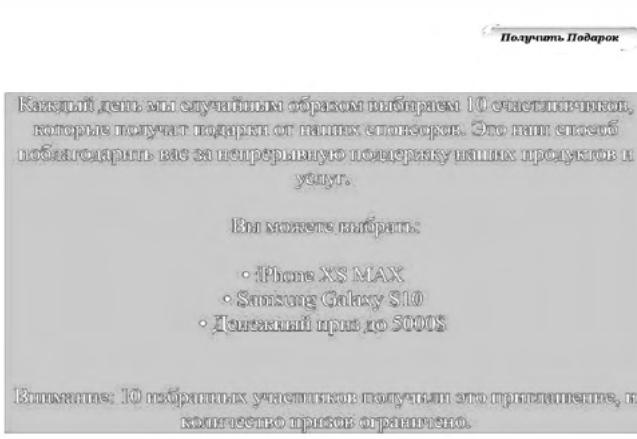
- Слова могут быть вытянуты и повернуты
- Плохое попадание в anchor boxes или receptive field



Детектирование текста

Второй путь: сегментировать текст и строить bbox по маске

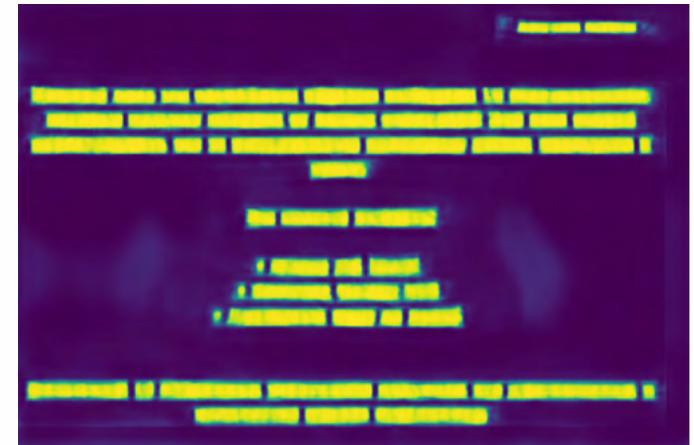
Примеры: FPN, U-net



Вопрос

Какой тип сегментации?

1. Semantic
2. Instance
3. Panoptic



💬 Пишите в чат

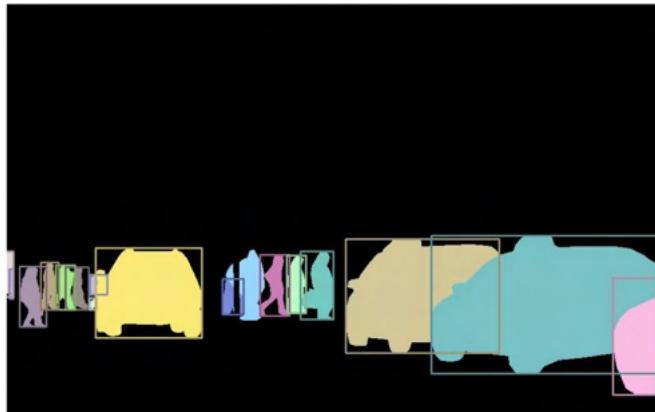
Детектирование текста



(a) image



(b) semantic segmentation



(c) instance segmentation



(d) panoptic segmentation

Детектирование текста

Трудности semantic segmentation:

- Соседние слова склеиваются
- Как объединять в строки?



Детектирование текста

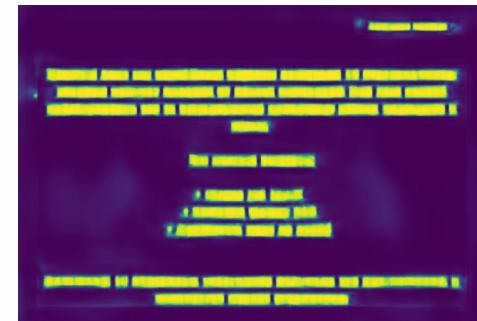
Трудности semantic segmentation:

- Соседние слова склеиваются
- Как объединять в строки?

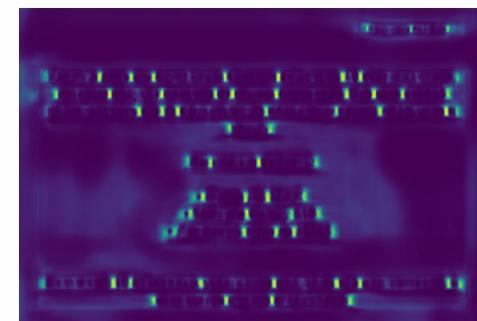
Решение:

- Добавить класс “разделитель”
- Добавить класс “строка”

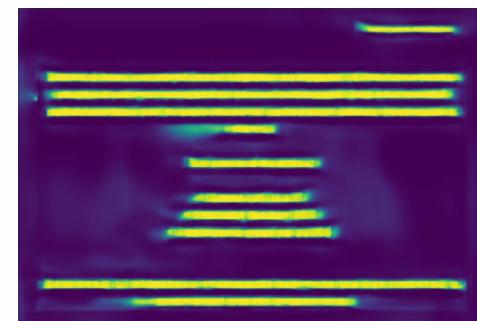
Слово



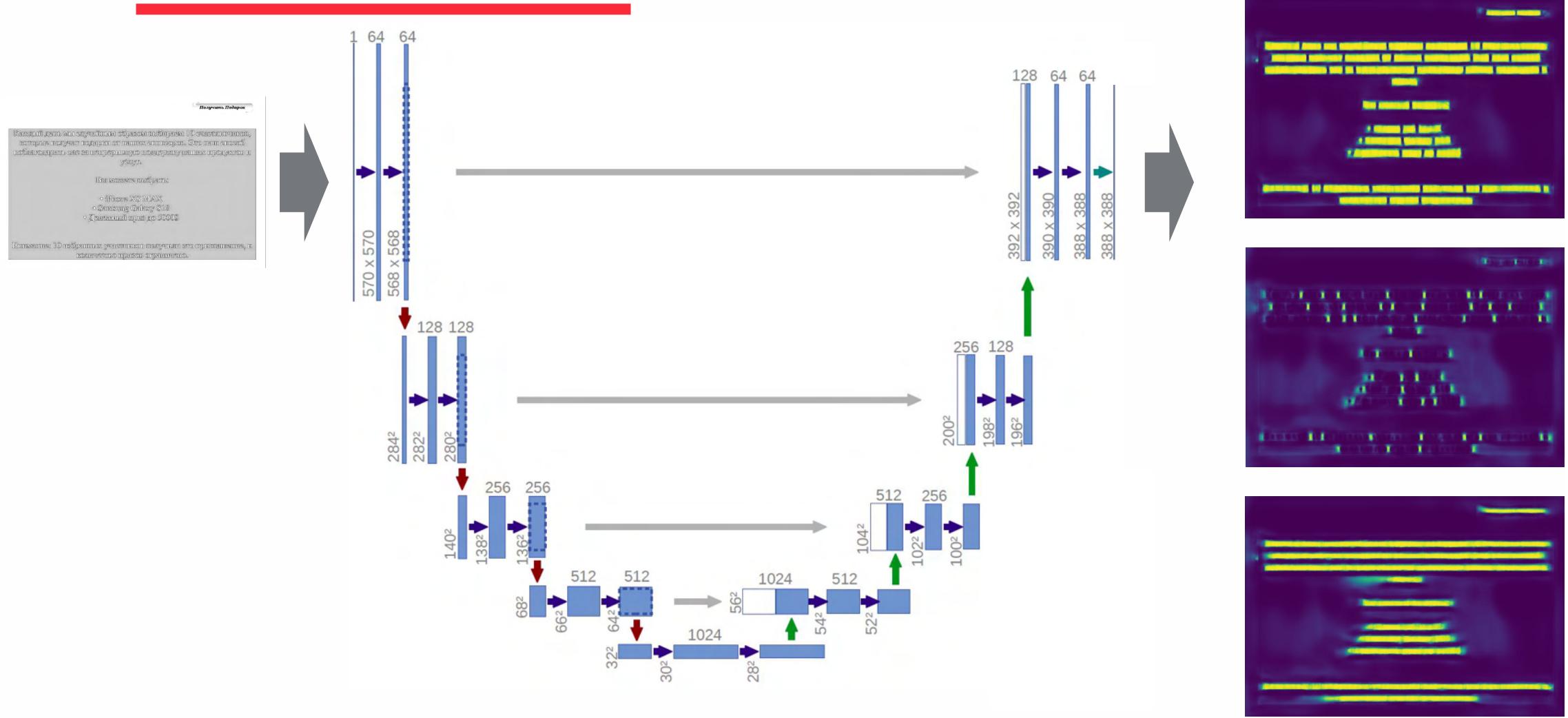
Разделитель



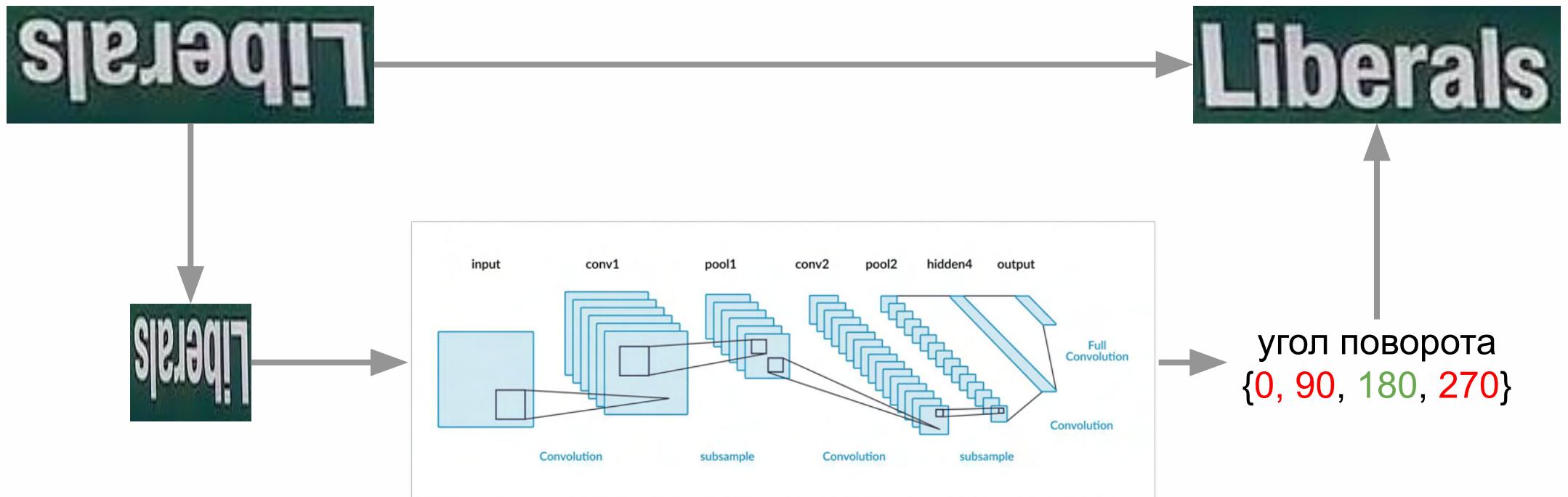
Строка



U-Net



Выравнивание ориентации



Вопросы

Распознавание

Вход распознавания



Liberals
our

Детектор

I-40

Нормализация

- Разный размер изображений
- Нужно составить батч



Пример распознавания



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
o	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
r	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
u	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0



T	o	T	u	u	T	r	T
---	---	---	---	---	---	---	---



"our"

Пример распознавания



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
o	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
r	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
u	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0



T	o	T	u	u	T	r	T
---	---	---	---	---	---	---	---

Зачем символ Т ?



Пишите в чат, говорите голосом



“our”

Пример распознавания



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
o	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
r	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
u	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Зачем символ Т ?

Пример: “apple”



T	o	T	u	u	T	r	T
---	---	---	---	---	---	---	---



“our”

Пример распознавания



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
o	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
r	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
u	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0



T	o	T	u	u	T	r	T
---	---	---	---	---	---	---	---



Небольшое упрощение:

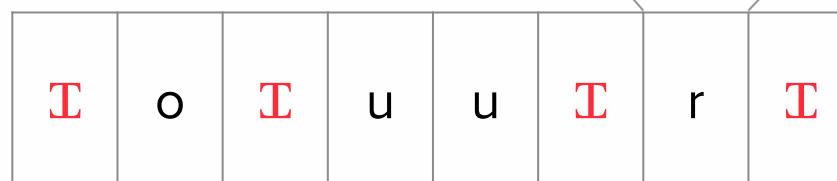
- Т в начале и конце
- Т обязательно между буквами

“our”

Свёрточная сеть



CNN receptive field



Контекст ограничен!

Вопрос

Какая архитектура поможет учесть весь контекст?

 Пишите в чат



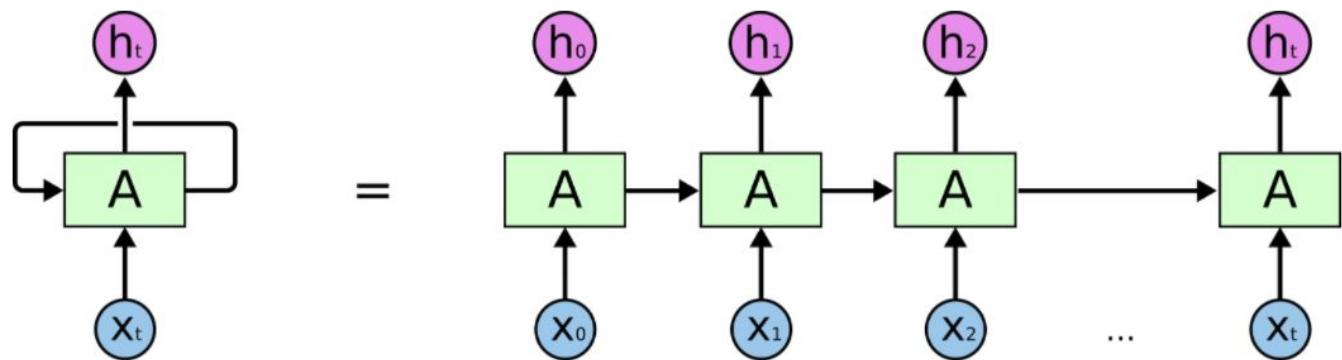
T	o	T	u	u	T	r	T
---	---	---	---	---	---	---	---

Задание

Тест про RNN

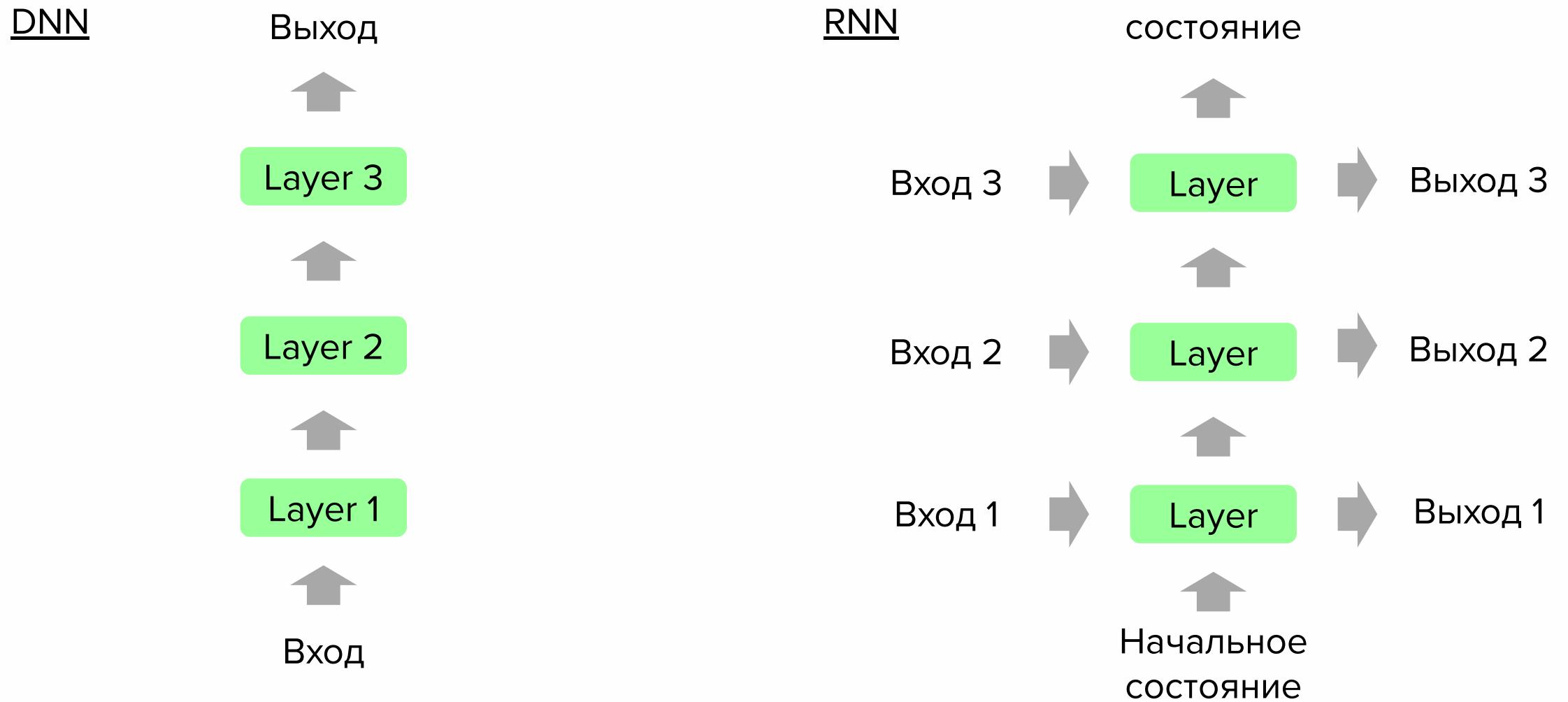
🔗 Ссылка в чате

⌚ 5 минут



An unrolled recurrent neural network.

DNN vs RNN





RNN

Начальное
состояние

x_1

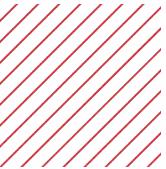
x_2

...

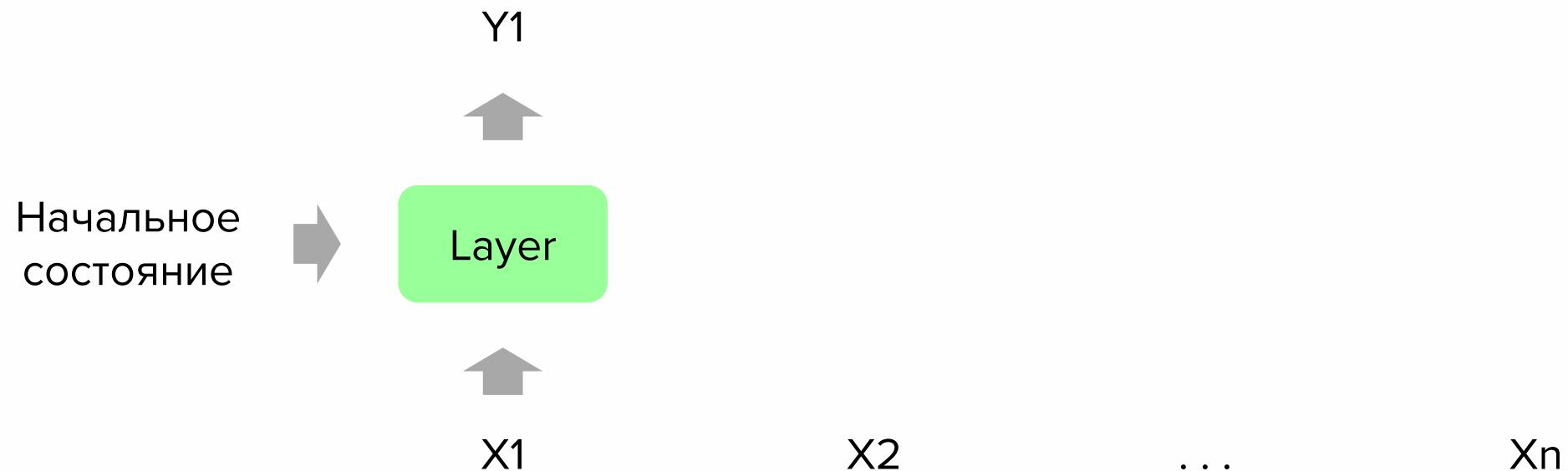
x_n

RNN

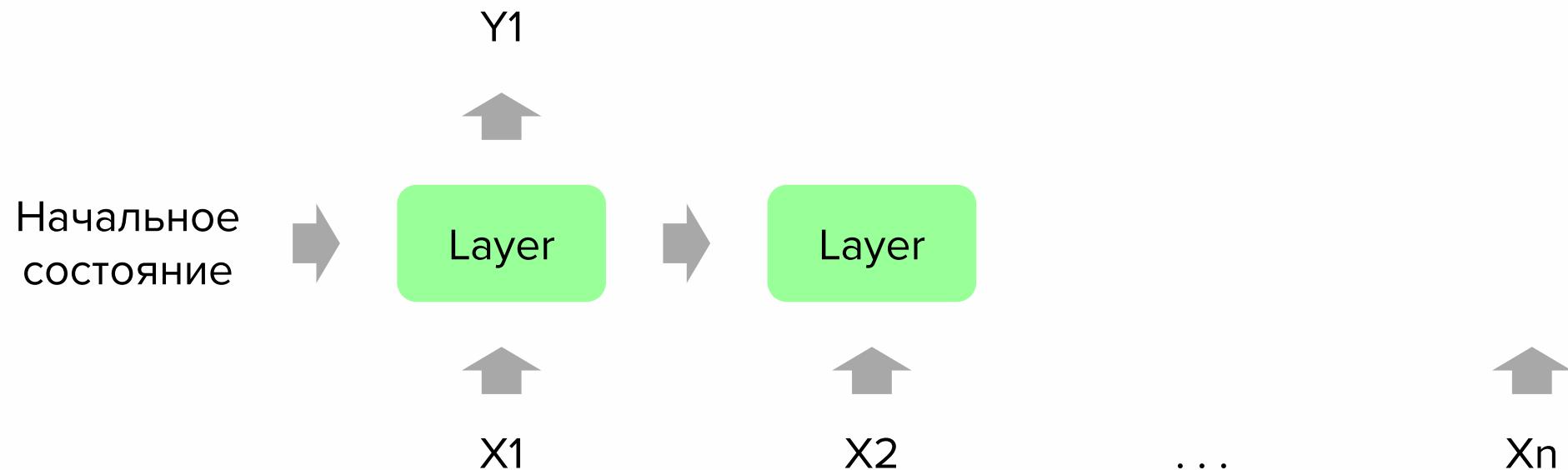




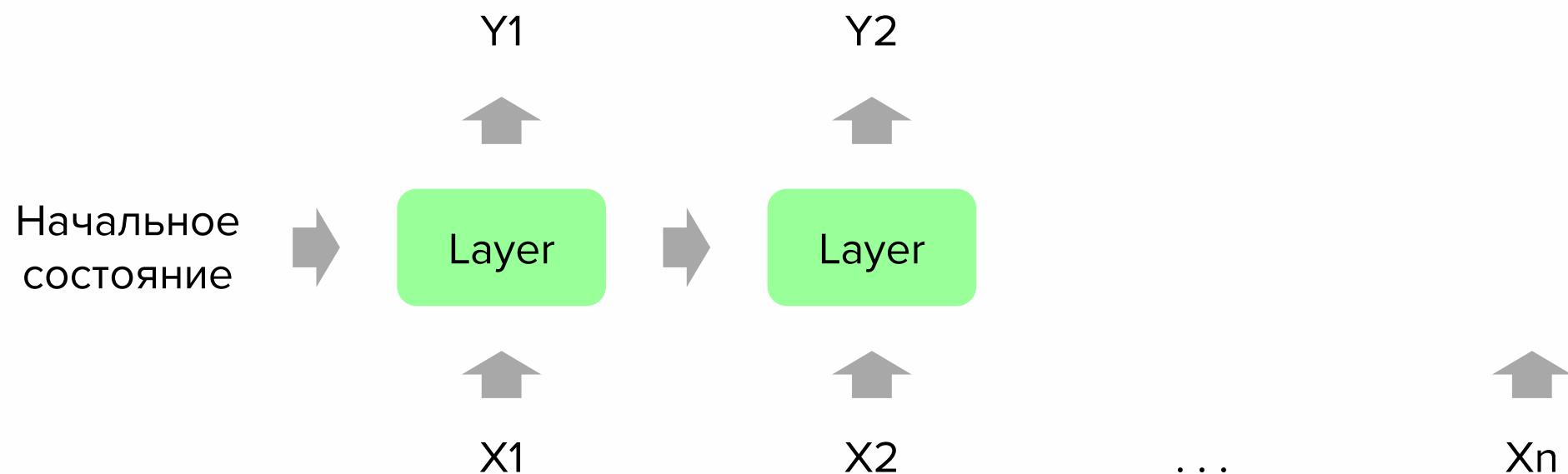
RNN



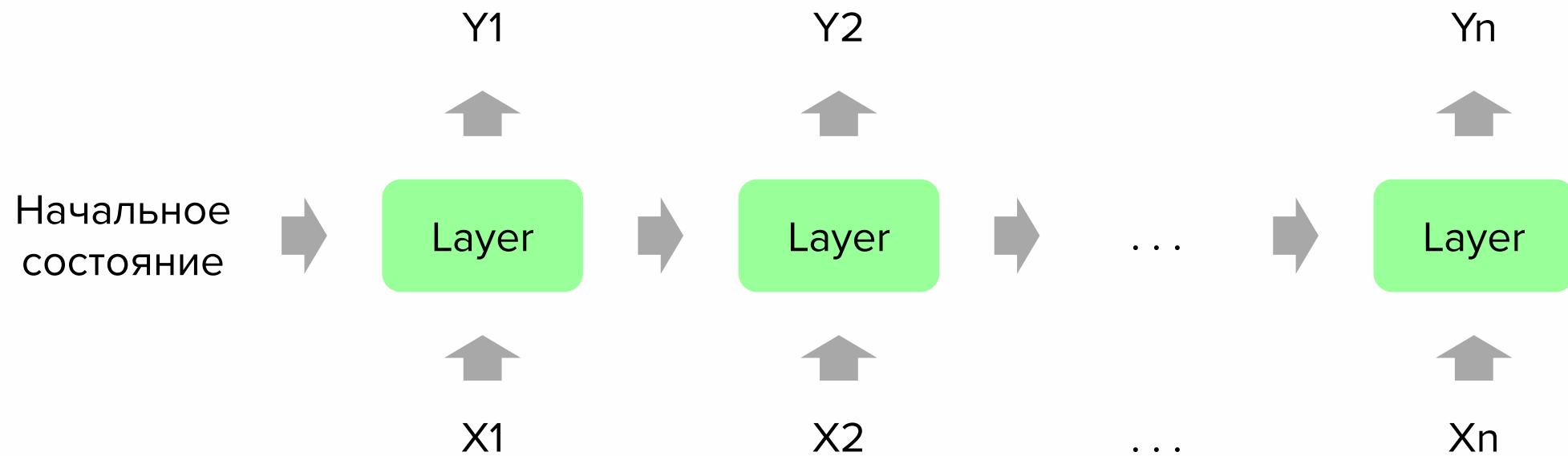
RNN



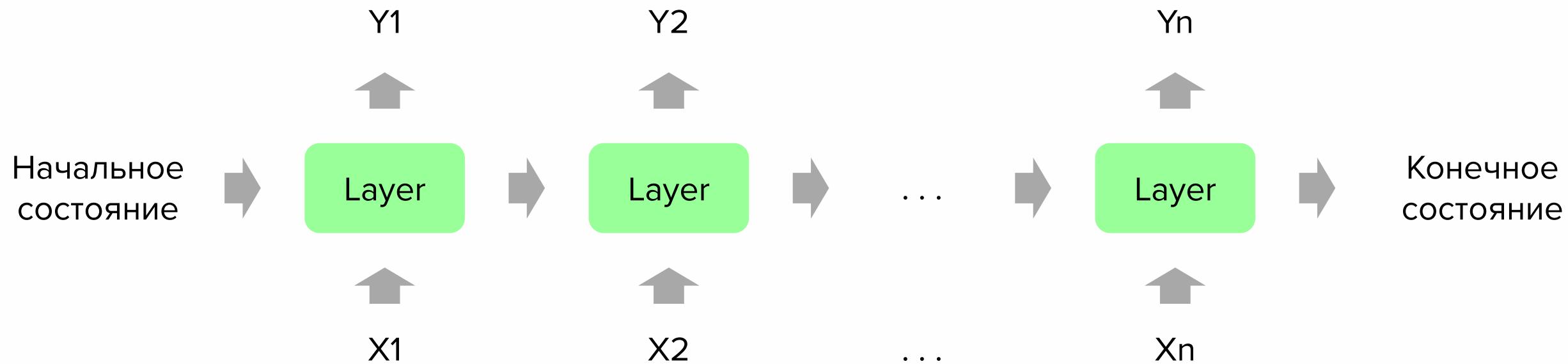
RNN



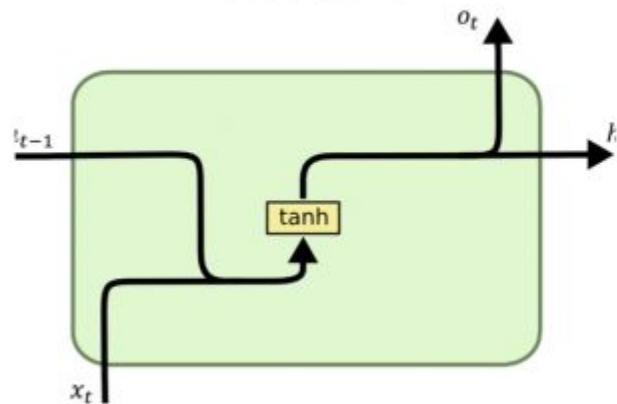
RNN



RNN



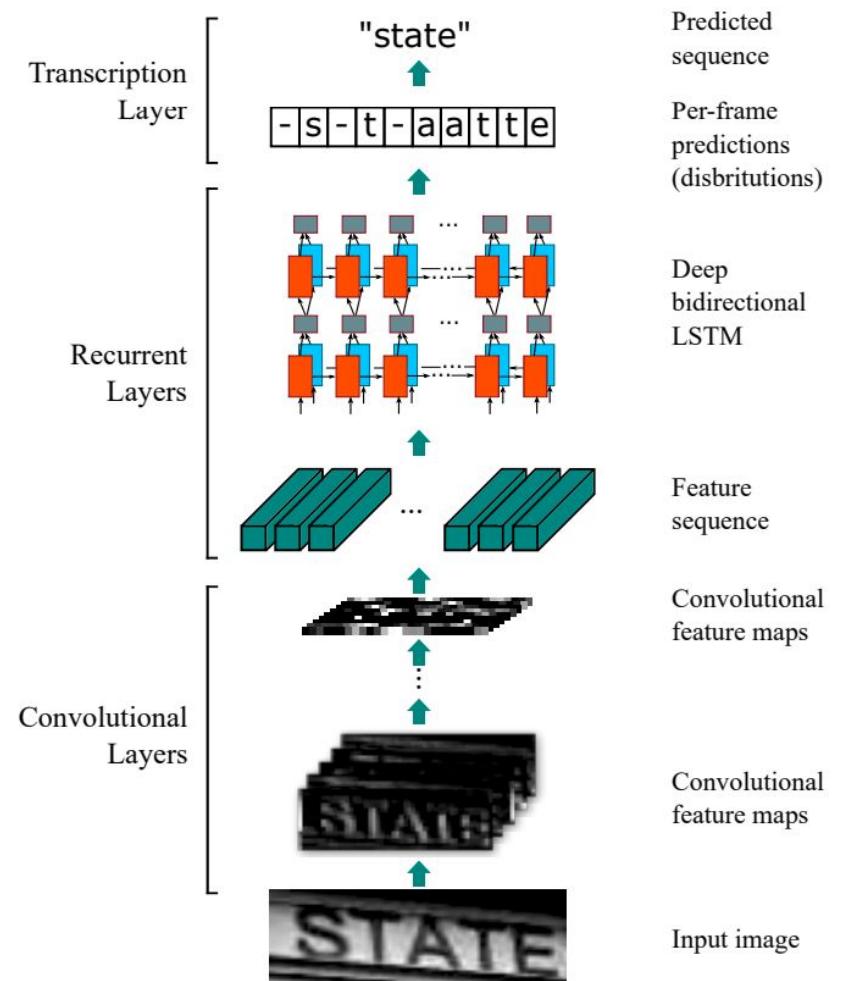
RNN



$$h_t = g(W_{hh}h_{t-1} + W_{xh}x_t + b_h)$$
$$o_t = h_t$$

Вопросы

CRNN = CNN + RNN



Пример распознавания



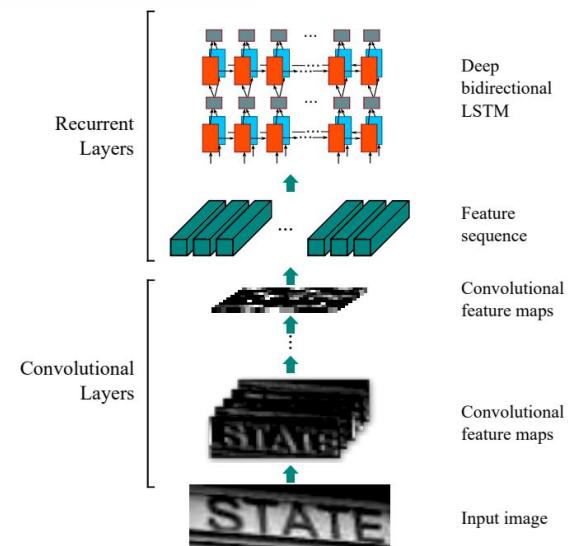
T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
r	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
u	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0



T	o	T	u	u	T	r	T
---	---	---	---	---	---	---	---



"our"



Перерыв

После поговорим про обучение

CTC loss

Разметка



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
o	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
r	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
u	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Небольшое упрощение:

- Т в начале и конце
- Т обязательно между буквами

Выравнивание

Метка

T	o	T	u	u	T	r	T
---	---	---	---	---	---	---	---



“our”

Вероятность выравнивания



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
r	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
u	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

T	O	T	u	u	T	r	T
---	---	---	---	---	---	---	---

0.9	0.8	0.8	0.8	0.6	0.7	0.7	0.6
-----	-----	-----	-----	-----	-----	-----	-----

$$\prod = 0.081$$

Вероятность слова



Т	о	Т	у	Т	р	р	Т
Т	о	Т	у	у	Т	р	Т
Т	о	о	Т	у	Т	р	Т

$$\prod = 0.081$$

$$\prod = 0.072$$

$$\prod = 0.005$$

$$\sum = 0.158$$

CTC Loss

Cross entropy / NLL

$$L(pred, label) = -\log P(label|pred)$$



¶	o	¶	u	¶	r	r	¶
¶	o	¶	u	u	¶	r	¶
¶	o	o	¶	u	¶	r	¶

“our”

Перебор выравниваний

s - число букв

l - длина выхода

N - число вариантов выравнивания

$$N = C_{l-s-1}^{s-1} = \frac{(l-s-1)!}{(l-2s)!(s-1)!}$$



$$s = 3, l = 8 \Rightarrow N = 3$$

$$s = 5, l = 32 \Rightarrow N = 14950$$



т	о	т	у	т	р	р	т
т	о	т	у	у	т	р	т
т	о	о	т	у	т	р	т

“our”

Вопросы

Вычисление CTC loss

В PyTorch уже реализован)

Обозначения

Выход модели

$$p_t(c)$$



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
R	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
U	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0
	t1	t2	t3	t4	t5	t6	t7	t8



“Время”

Обозначения

Выход модели

$$p_t(c)$$



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
R	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
U	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0
	t1	t2	t3	t4	t5	t6	t7	t8

$$p_3(o) = 0.1$$

“Время”

Обозначения

Выход модели

$$p_t(c)$$



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
R	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
U	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0
	t1	t2	t3	t4	t5	t6	t7	t8

$$p_6(I) = 0.7$$

“Время”

Обозначения

Состояния

“our”



т	о	т	у	т	р	т
---	---	---	---	---	---	---

s1 s2 s3 s4 s5 s6 s7



Динамическое программирование

1

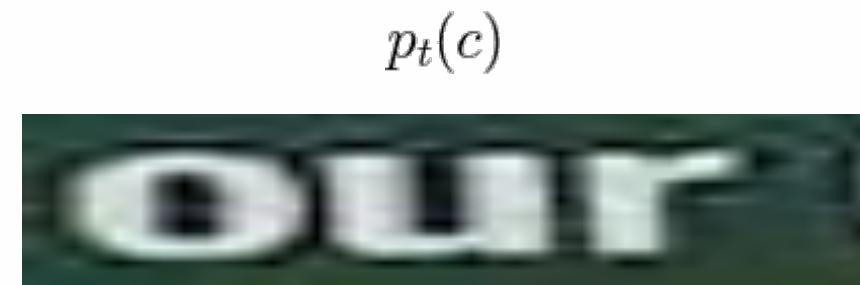
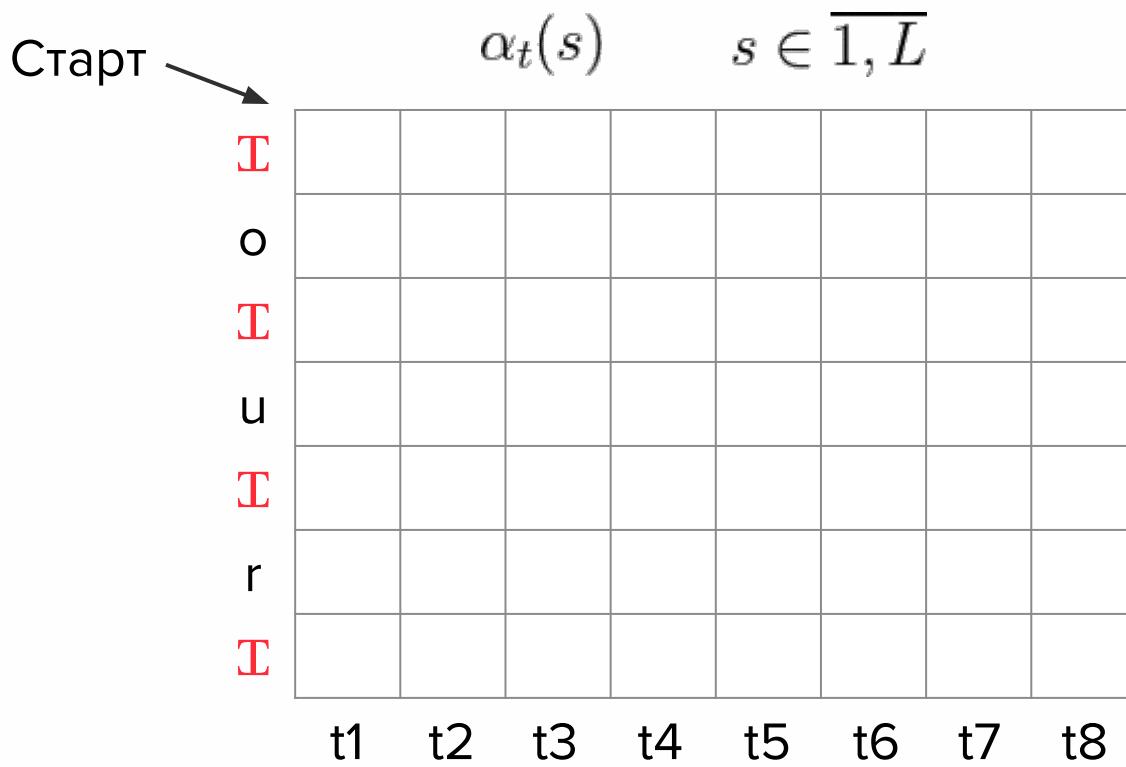
Разбить на подзадачи

2

Использовать ответы простых подзадач при решении сложных

Прямой проход

$\alpha_t(s)$ - вероятность префикса длины s в момент t



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
U	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
R	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Прямой проход

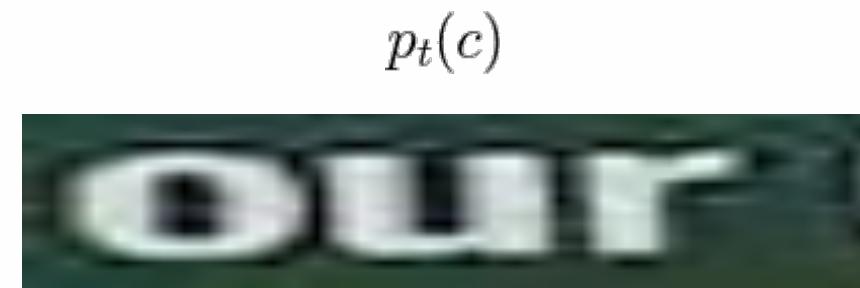
$\alpha_t(s)$ - вероятность префикса длины s в момент t

Старт →

$\alpha_t(s) \quad s \in \overline{1, L}$

T	0.9							
O								
T								
U								
E								
R								
T								
	t1	t2	t3	t4	t5	t6	t7	t8

Финиш →



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
R	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
U	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Прямой проход

$\alpha_t(s)$ - вероятность префикса длины s в момент t

Старт $\alpha_t(s) = (\alpha_{t-1}(s) + \alpha_{t-1}(s - 1))p_t(c(s))$

T	0.9	0.18						
O								
I								
U								
R								
T								
	t1	t2	t3	t4	t5	t6	t7	t8

Финиш

$$p_t(c)$$



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
I	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
U	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Прямой проход

$\alpha_t(s)$ - вероятность префикса длины s в момент t

Старт $\alpha_t(s) = (\alpha_{t-1}(s) + \alpha_{t-1}(s - 1))p_t(c(s))$

T	0.9	0.18						
O		0.72						
I								
U								
R								
T								
	t1	t2	t3	t4	t5	t6	t7	t8

$p_t(c)$



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
R	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
U	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Финиш

Прямой проход

$\alpha_t(s)$ - вероятность префикса длины s в момент t

Старт $\alpha_t(s) = (\alpha_{t-1}(s) + \alpha_{t-1}(s - 1))p_t(c(s))$

T	0.9	0.18						
O		0.72	0.09					
I								
U								
R								
T								
	t1	t2	t3	t4	t5	t6	t7	t8

$p_t(c)$



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
r	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
u	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Финиш

Задание

💡 Пишите в чат

$\alpha_t(s)$ - вероятность префикса длины s в момент t

Старт $\alpha_t(s) = (\alpha_{t-1}(s) + \alpha_{t-1}(s - 1))p_t(c(s))$

T	0.9	0.18	?					
O		0.72	0.09					
I								
U								
R								
T								
	t1	t2	t3	t4	t5	t6	t7	t8

$p_t(c)$



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
I	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
U	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Финиш

Прямой проход

$\alpha_t(s)$ - вероятность префикса длины s в момент t

Старт $\alpha_t(s) = (\alpha_{t-1}(s) + \alpha_{t-1}(s - 1))p_t(c(s))$

T	0.9	0.18	0.144					
O		0.72	0.09					
I								
U								
R								
T								
	t1	t2	t3	t4	t5	t6	t7	t8

$p_t(c)$



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
I	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
U	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Финиш

Прямой проход

$\alpha_t(s)$ - вероятность префикса длины s в момент t

Старт $\alpha_t(s) = (\alpha_{t-1}(s) + \alpha_{t-1}(s - 1))p_t(c(s))$

T	0.9	0.18	0.144					
O	0	0.72	0.09					
T	0	0						
U								
R								
T								
	t1	t2	t3	t4	t5	t6	t7	t8

$p_t(c)$



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
R	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
U	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Финиш

Задание

💡 Пишите в чат

$\alpha_t(s)$ - вероятность префикса длины s в момент t

Старт $\alpha_t(s) = (\alpha_{t-1}(s) + \alpha_{t-1}(s - 1))p_t(c(s))$

T	0.9	0.18	0.144					
O	0	0.72	0.09					
I	0	0	?					
U								
R								
T								
	t1	t2	t3	t4	t5	t6	t7	t8

$p_t(c)$



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
I	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
U	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Финиш

Прямой проход

$\alpha_t(s)$ - вероятность префикса длины s в момент t

Старт $\alpha_t(s) = (\alpha_{t-1}(s) + \alpha_{t-1}(s - 1))p_t(c(s))$

T	0.9	0.18	0.144					
O	0	0.72	0.09					
I	0	0	0.576					
U								
R								
T								

t1 t2 t3 t4 t5 t6 t7 t8

Финиш

$p_t(c)$

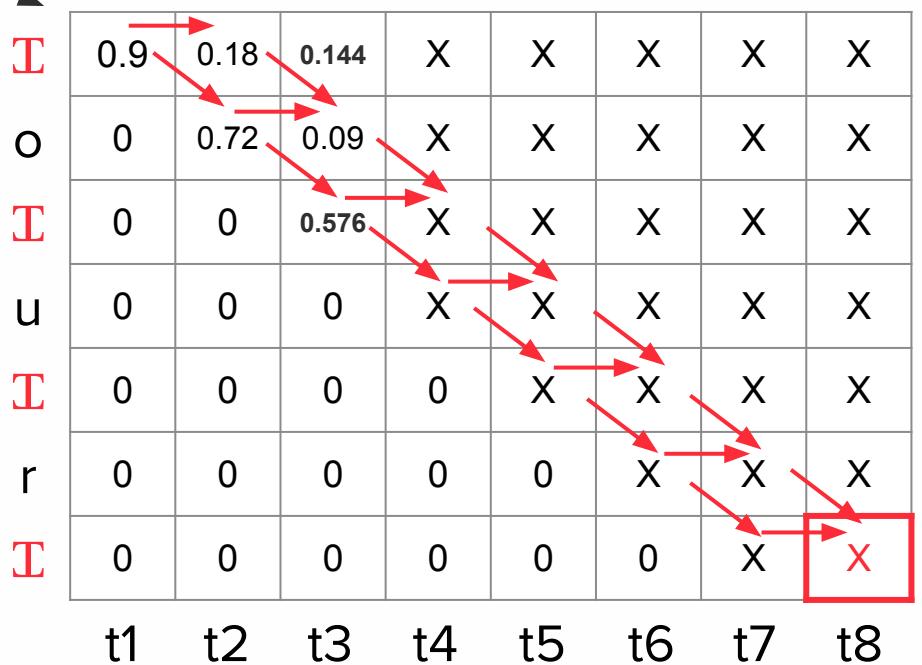


T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
I	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
U	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Прямой проход

$\alpha_t(s)$ - вероятность префикса длины s в момент t

Старт $\alpha_t(s) = (\alpha_{t-1}(s) + \alpha_{t-1}(s - 1))p_t(c(s))$



$p_t(c)$

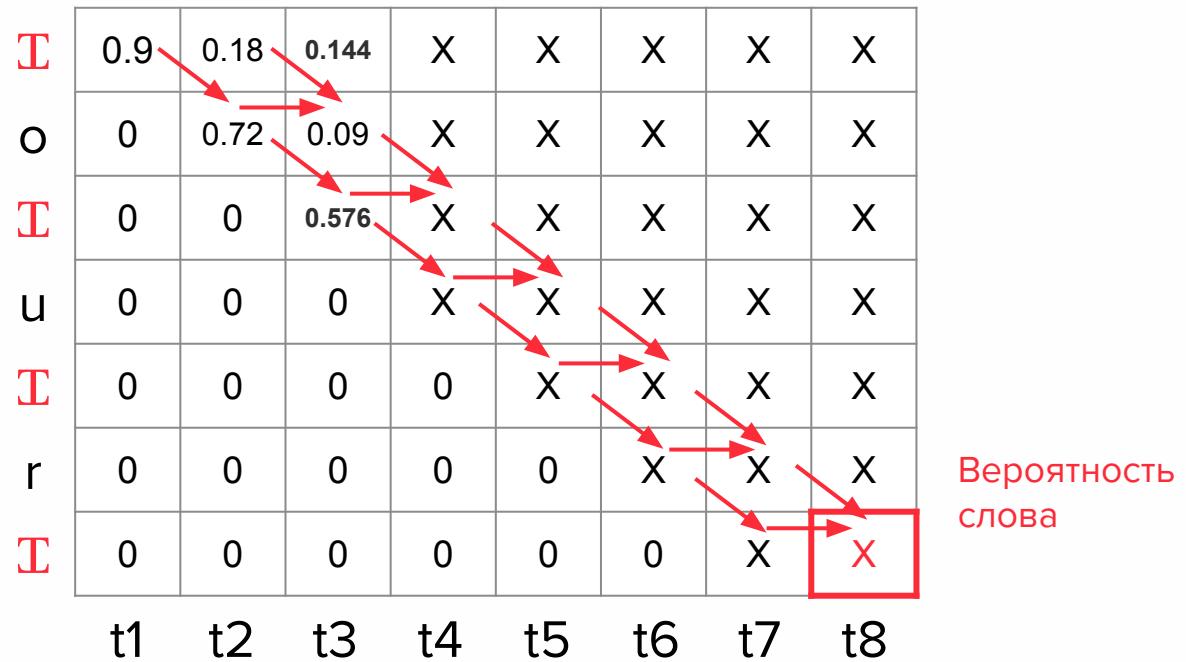


Т	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
О	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
У	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
Р	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Вопросы

Производная

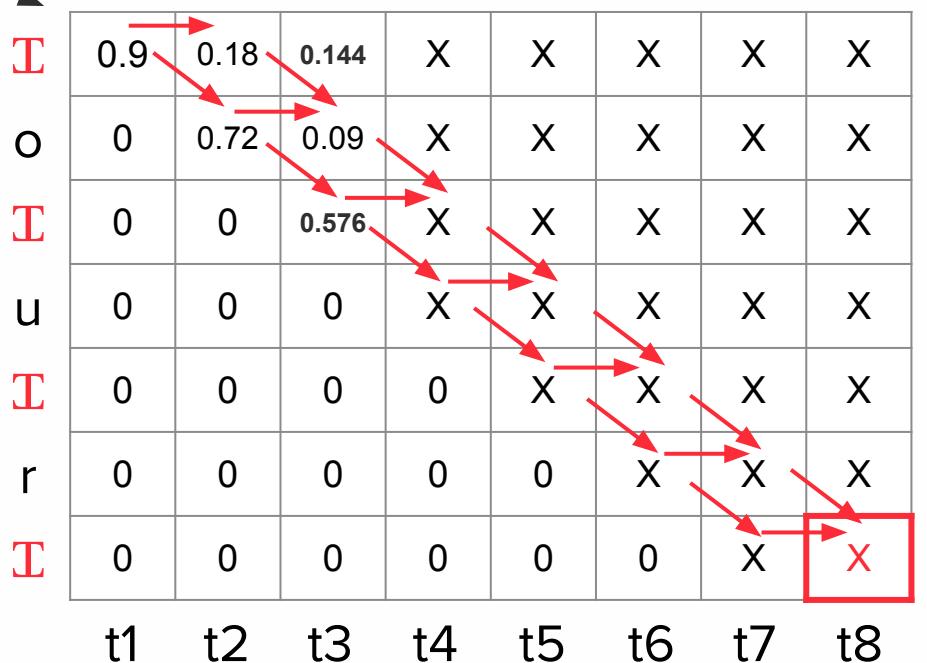
- Вероятность дифференцируема
 - Производная вычисляется шаг за шагом с конца
 - Не параллелистся



Как выглядел прямой проход

$\alpha_t(s)$ - вероятность префикса длины s в момент t

Старт $\alpha_t(s) = (\alpha_{t-1}(s) + \alpha_{t-1}(s - 1))p_t(c(s))$



$p_t(c)$

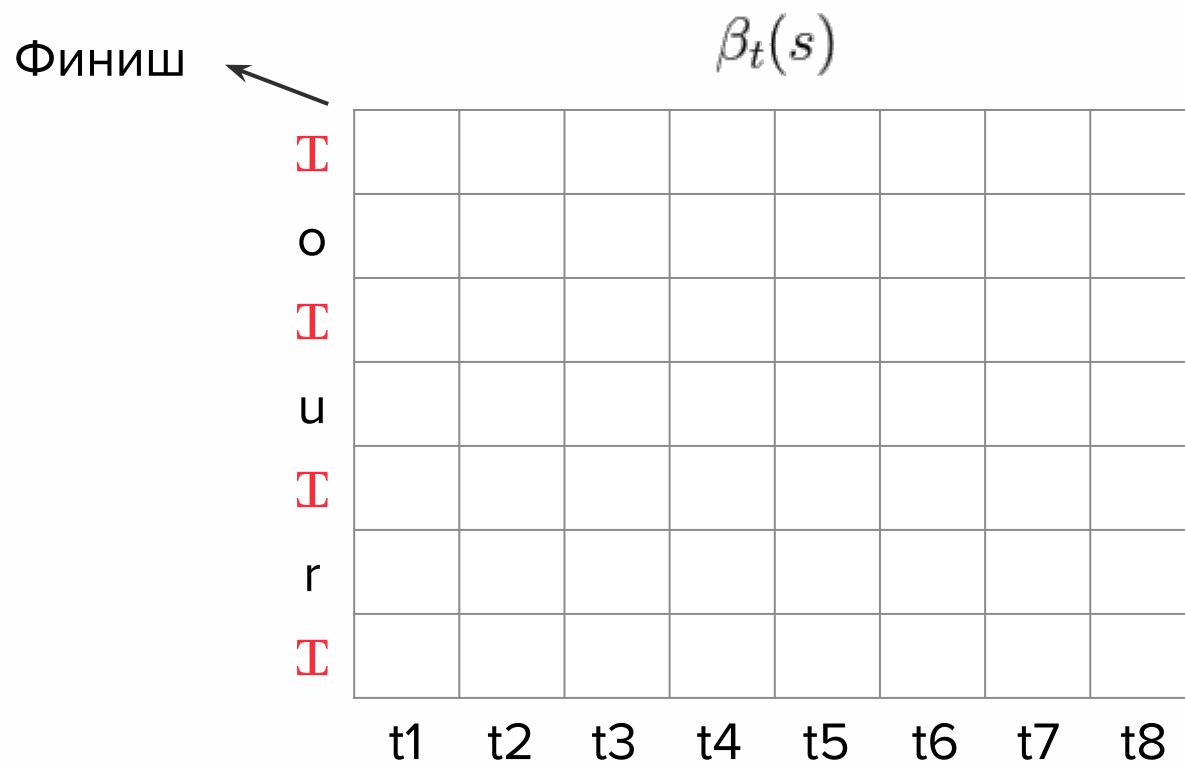


T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
I	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
U	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Финиш

Обратный проход

$\beta_t(s)$ - вероятность суффикса $s : L$, начиная с момента t

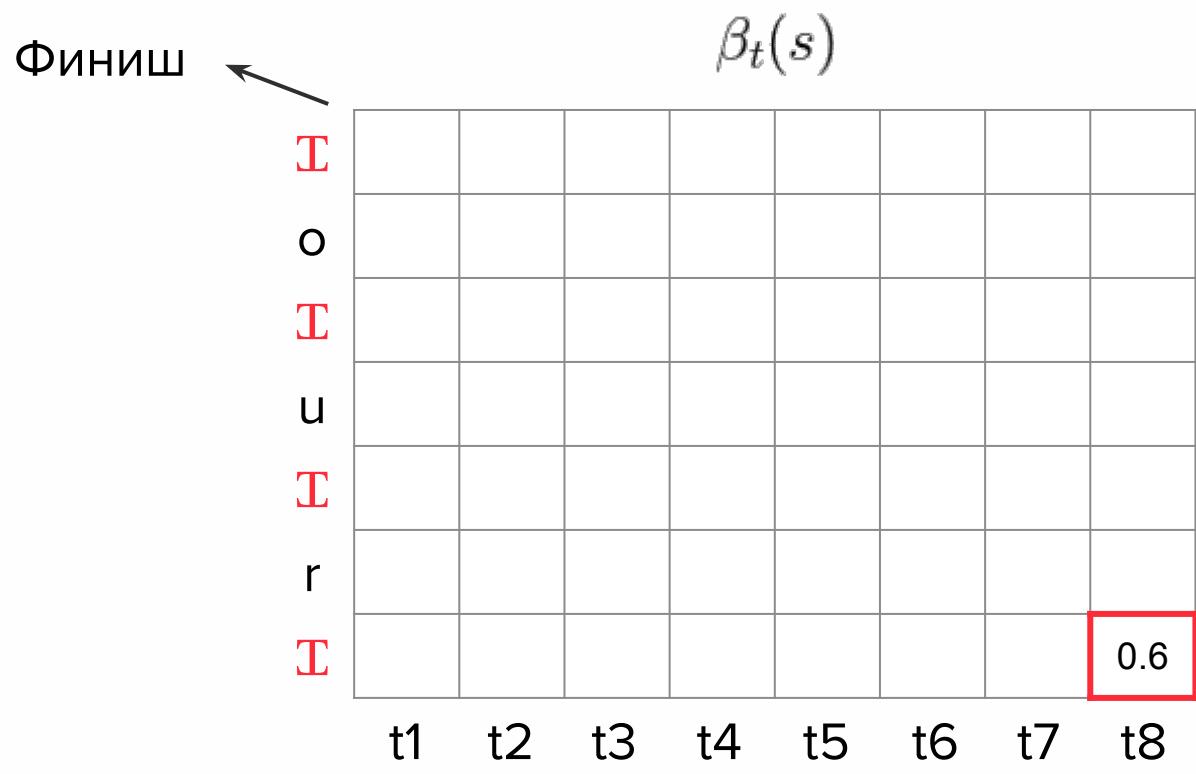


$p_t(c)$

Т	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
о	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
у	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
р	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Обратный проход

$\beta_t(s)$ - вероятность суффикса $s : L$, начиная с момента t



$p_t(c)$

T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
I	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
U	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Обратный проход

$\beta_t(s)$ - вероятность суффикса $s : L$, начиная с момента t

Финиш

$\beta_t(s)$

Т
о
т
у
т
р
т

t1 t2 t3 t4 t5 t6 t7 t8

0.18 0.6

Старт

$p_t(c)$



Т
о
т
у

0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Обратный проход

$\beta_t(s)$ - вероятность суффикса $s : L$, начиная с момента t

Финиш

$\beta_t(s)$

Т
о
т
у
т
р
т

t1 t2 t3 t4 t5 t6 t7 t8

0.42
0.18 0.6

Старт

$p_t(c)$



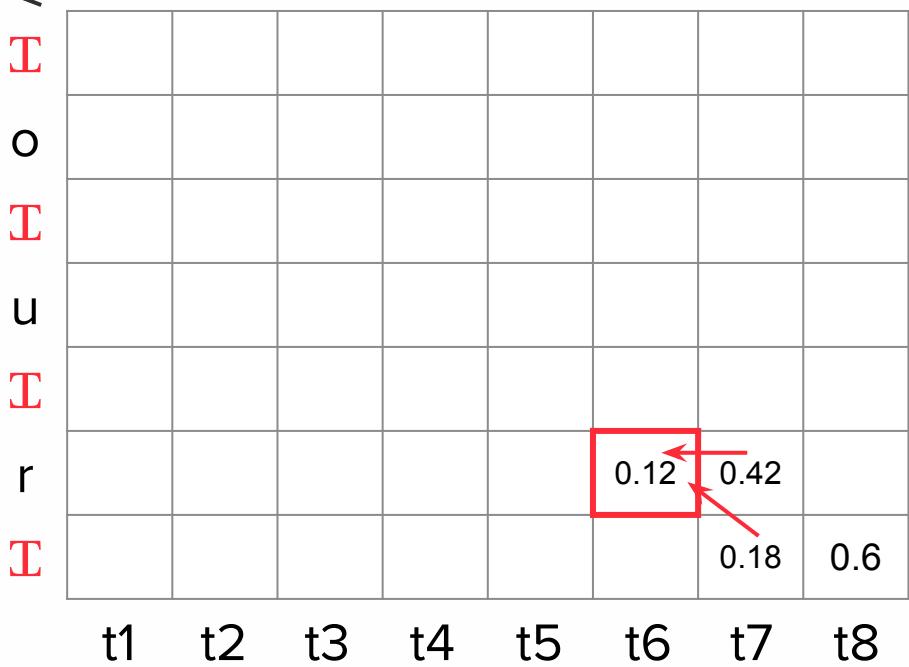
Т
о
т
у
т

0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Обратный проход

$\beta_t(s)$ - вероятность суффикса $s : L$, начиная с момента t

Финиш $\leftarrow \beta_t(s) = (\beta_{t+1}(s + 1) + \beta_{t+1}(s))p_t(c(s))$



$$p_t(c)$$



Т	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
О	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
У	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
Р	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Обратный проход

$\beta_t(s)$ - вероятность суффикса $s : L$, начиная с момента t

Финиш

$$\beta_t(s) = (\beta_{t+1}(s + 1) + \beta_{t+1}(s))p_t(c(s))$$

T	X	X	X	0	0	0	0	0
O	X	X	X	0	0	0	0	0
T	X	X	X	0	0	0	0	0
U	X	X	X	X	0	0	0	0
E	X	X	X	X	X	0	0	0
R	X	X	X	X	X	0.12	0.42	0
T	X	X	X	X	X	X	0.18	0.6
	t1	t2	t3	t4	t5	t6	t7	t8

Старт

$$p_t(c)$$



T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
R	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
U	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

Вопросы

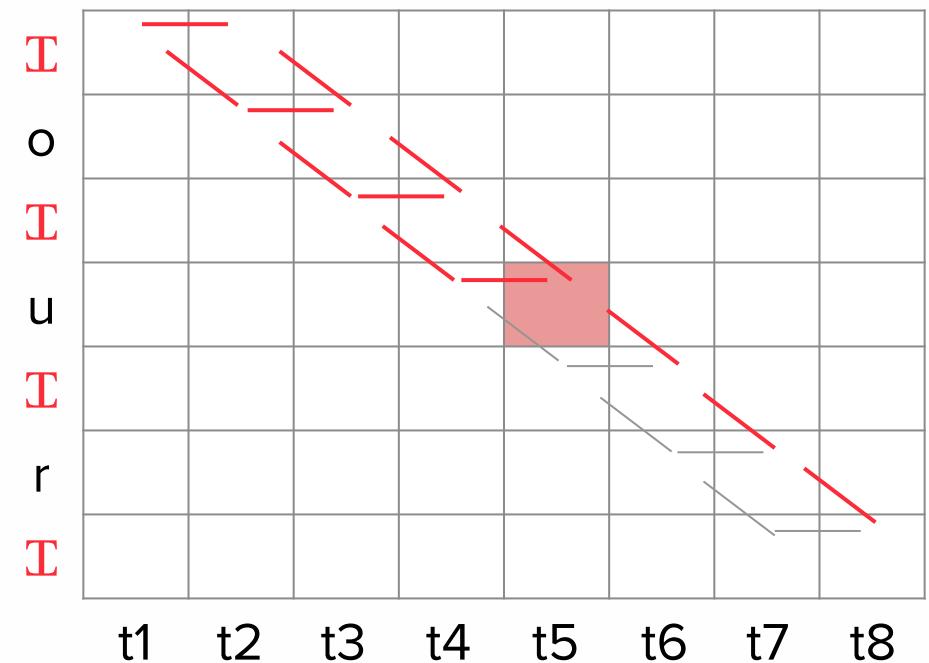
Magic

$\alpha_t(s)$ - вероятность префикса длины s в момент t

$\beta_t(s)$ - вероятность суффикса $s : T$, начиная с момента t

$$\gamma_t(s) = \frac{\alpha_t(s)\beta_t(s)}{p_t(c(s))}$$

$\gamma_t(s)$ - вероятность всех путей, проходящих (t, s)



Magic

$\alpha_t(s)$ - вероятность префикса длины s в момент t

$\beta_t(s)$ - вероятность суффикса $s : T$, начиная с момента t

$$\gamma_t(s) = \frac{\alpha_t(s)\beta_t(s)}{p_t(c(s))}$$

$\gamma_t(s)$ - вероятность всех путей, проходящих (t, s)



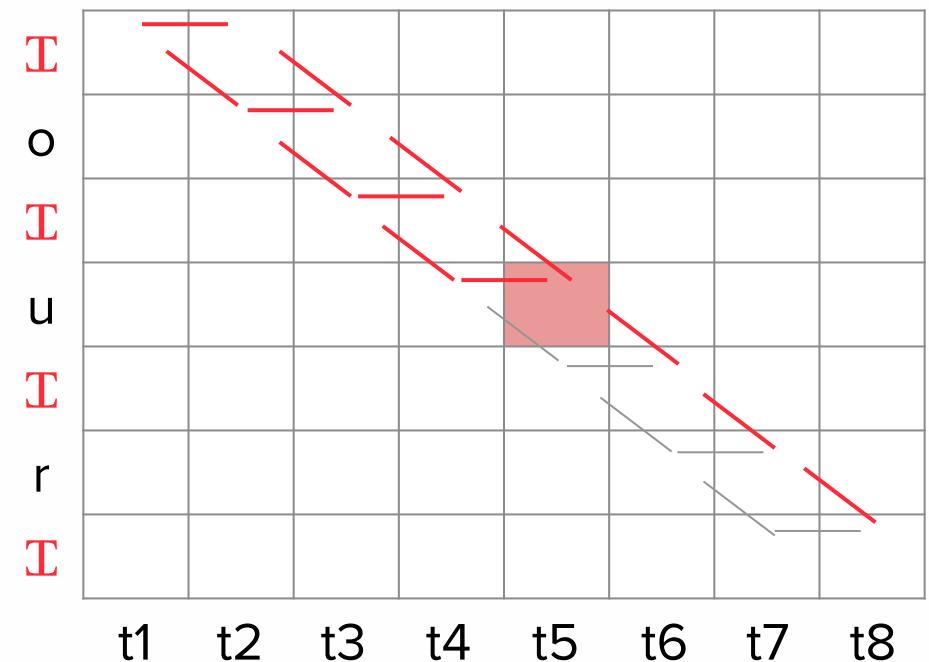
осознать



1 минута



Вопросы в чат или голосом

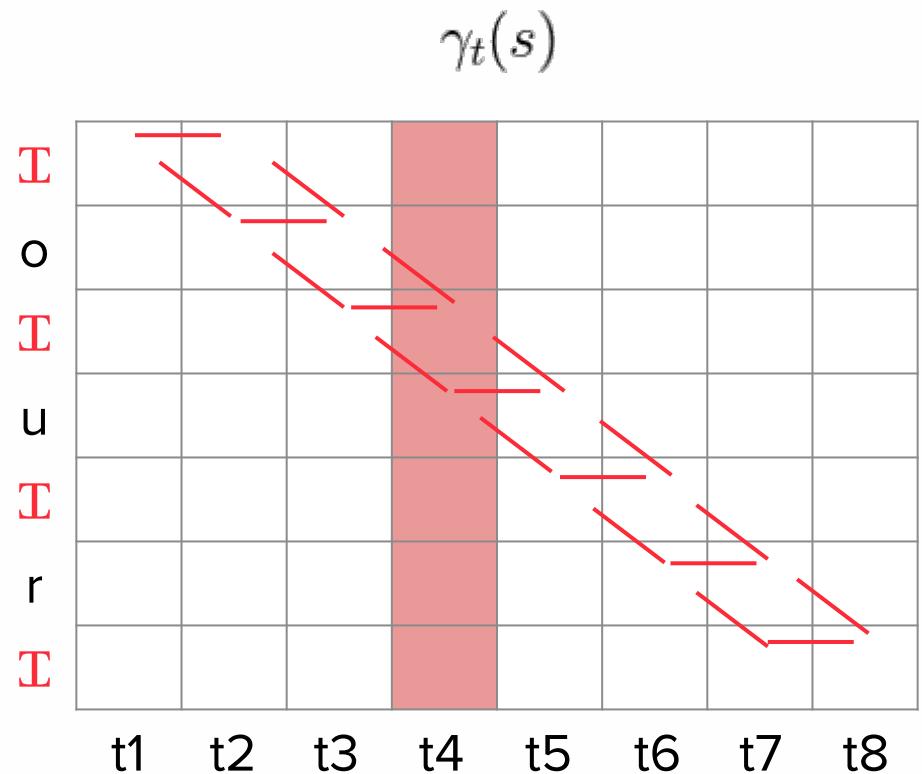


Magic

$\gamma_t(s)$ - вероятность всех путей, проходящих (t, s)

$$\gamma_t(s) = \frac{\alpha_t(s)\beta_t(s)}{p_t(c(s))}$$

$$P(label|pred) = \sum_s \gamma_t(s), \forall t$$



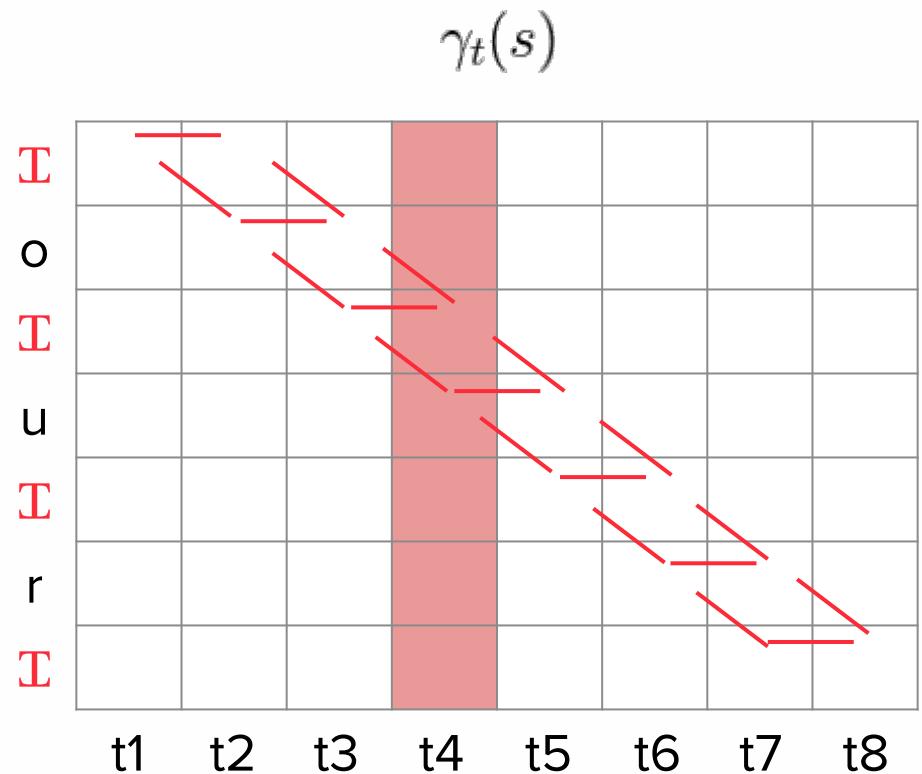
Magic

$\gamma_t(s)$ - вероятность всех путей, проходящих (t, s)

$$\gamma_t(s) = \frac{\alpha_t(s)\beta_t(s)}{p_t(c(s))}$$

$$P(label|pred) = \sum_s \gamma_t(s), \forall t$$

$$L(pred, label) = -\log P(label|pred)$$



Градиент СТС

$\alpha_t(s)$ - вероятность префикса длины s в момент t

$\beta_t(s)$ - вероятность суффикса $s : T$, начиная с момента t

$\gamma_t(s)$ - вероятность всех путей, проходящих (t, s)

$$\alpha_t(s) = (\alpha_{t-1}(s) + \alpha_{t-1}(s-1))p_t(c(s))$$

$$\beta_t(s) = (\beta_{t+1}(s+1) + \beta_{t+1}(s))p_t(c(s))$$

$$\gamma_t(s) = \frac{\alpha_t(s)\beta_t(s)}{p_t(c(s))}$$

$$\alpha_t(s)'_{p_t(c(s))} = \frac{\alpha_t(s)}{p_t(c(s))} \quad \beta_t(s)'_{p_t(c(s))} = \frac{\beta_t(s)}{p_t(c(s))}$$

$$\gamma_t(s)'_{p_t(c(s))} = \frac{\alpha_t(s)\beta_t(s)}{p_t^2(c(s))} = \frac{\gamma_t(s)}{p_t(c(s))}$$

Градиент СТС

$\alpha_t(s)$ - вероятность префикса длины s в момент t

$\beta_t(s)$ - вероятность суффикса $s : T$, начиная с момента t

$\gamma_t(s)$ - вероятность всех путей, проходящих (t, s)

$$\alpha_t(s) = (\alpha_{t-1}(s) + \alpha_{t-1}(s-1))p_t(c(s))$$

$$\beta_t(s) = (\beta_{t+1}(s+1) + \beta_{t+1}(s))p_t(c(s))$$

$$\gamma_t(s) = \frac{\alpha_t(s)\beta_t(s)}{p_t(c(s))}$$

$$\alpha_t(s)'_{p_t(c(s))} = \frac{\alpha_t(s)}{p_t(c(s))} \quad \beta_t(s)'_{p_t(c(s))} = \frac{\beta_t(s)}{p_t(c(s))}$$

$$\gamma_t(s)'_{p_t(c(s))} = \frac{\alpha_t(s)\beta_t(s)}{p_t^2(c(s))} = \frac{\gamma_t(s)}{p_t(c(s))}$$



осознать



1 минута



Вопросы в чат или голосом

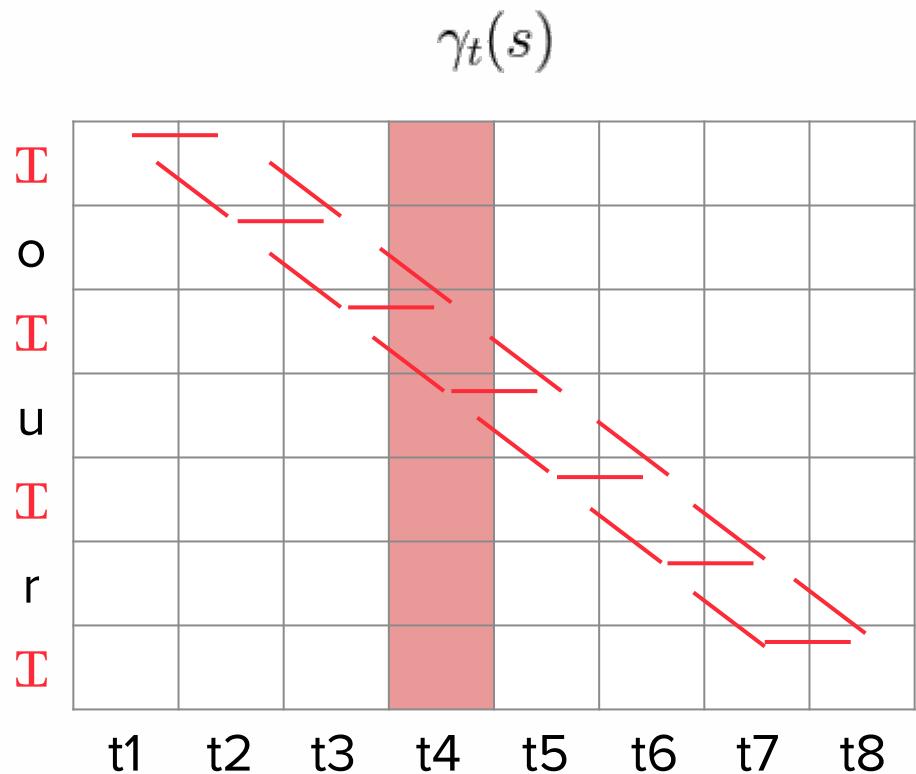
Градиент СТС

$\gamma_t(s)$ - вероятность всех путей, проходящих (t, s)

Уже
знаем

$$P(label|pred) = \sum_s \gamma_t(s), \forall t$$

$$L(pred, label) = -\log P(label|pred)$$



Градиент СТС

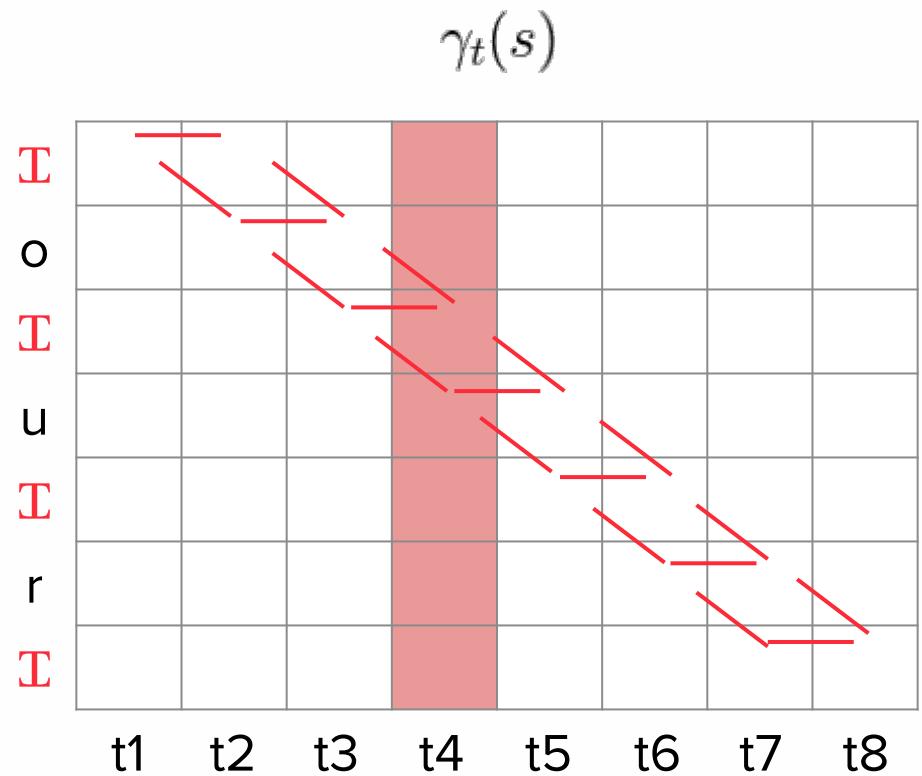
$\gamma_t(s)$ - вероятность всех путей, проходящих (t, s)

$$P(label|pred) = \sum_s \gamma_t(s), \forall t$$

$$L(pred, label) = -\log P(label|pred)$$

$$L'_{p_t(c)} = -\frac{1}{P(label|pred)} \sum_{s:c(s)=c} \gamma_t(s)'_{p_t(c)}$$

$$L'_{p_t(c)} = -\frac{1}{P(label|pred)} \sum_{s:c(s)=c} \frac{\gamma_t(s)}{p_t(c)}$$

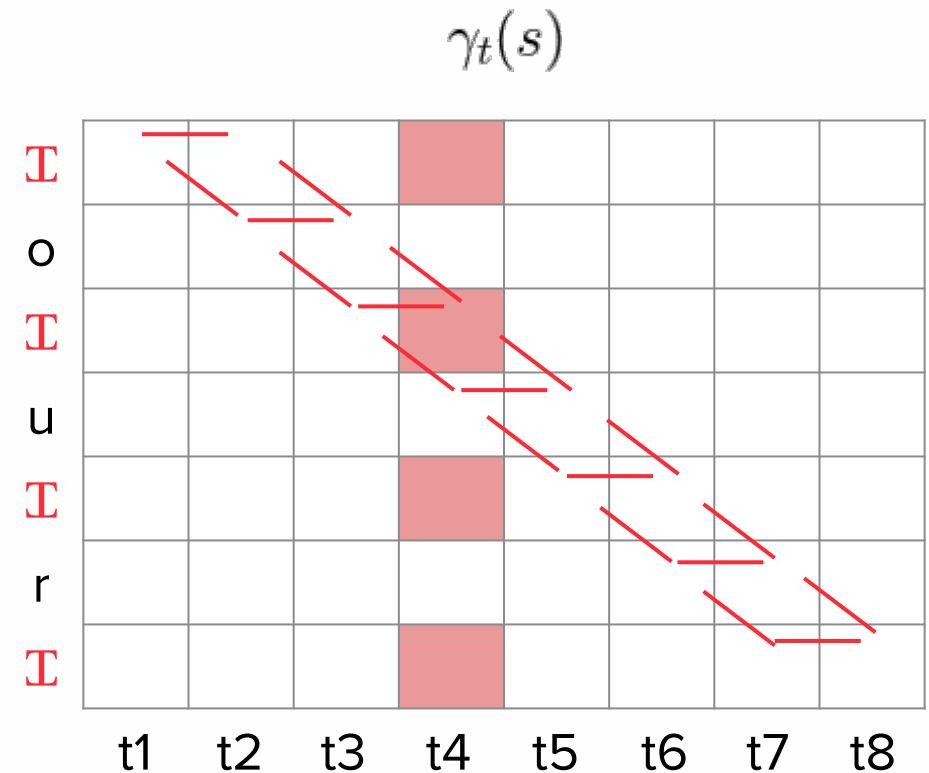


Градиент СТС

$\gamma_t(s)$ - вероятность всех путей, проходящих (t, s)

$$L'_{p_4(\text{I})} = -\frac{1}{P(\text{label}|\text{pred})} \sum_{s:c(s)=\text{I}} \frac{\gamma_4(s)}{p_4(\text{I})}$$

	pred							
T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
r	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
u	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0



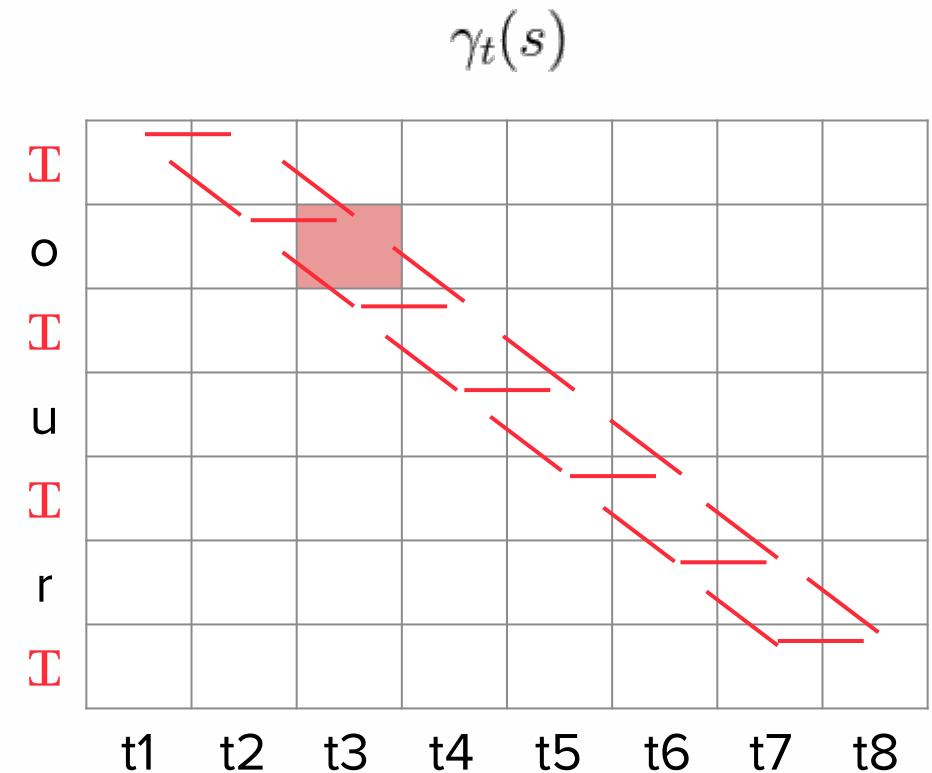
Градиент СТС

$\gamma_t(s)$ - вероятность всех путей, проходящих (t, s)

$$L'_{p_3(o)} = -\frac{1}{P(label|pred)} \sum_{s:c(s)=o} \frac{\gamma_3(s)}{p_3(o)}$$

	T	O	R	P	E	S	A	N
T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
R	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
P	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0
E								
S								
A								
N								

pred



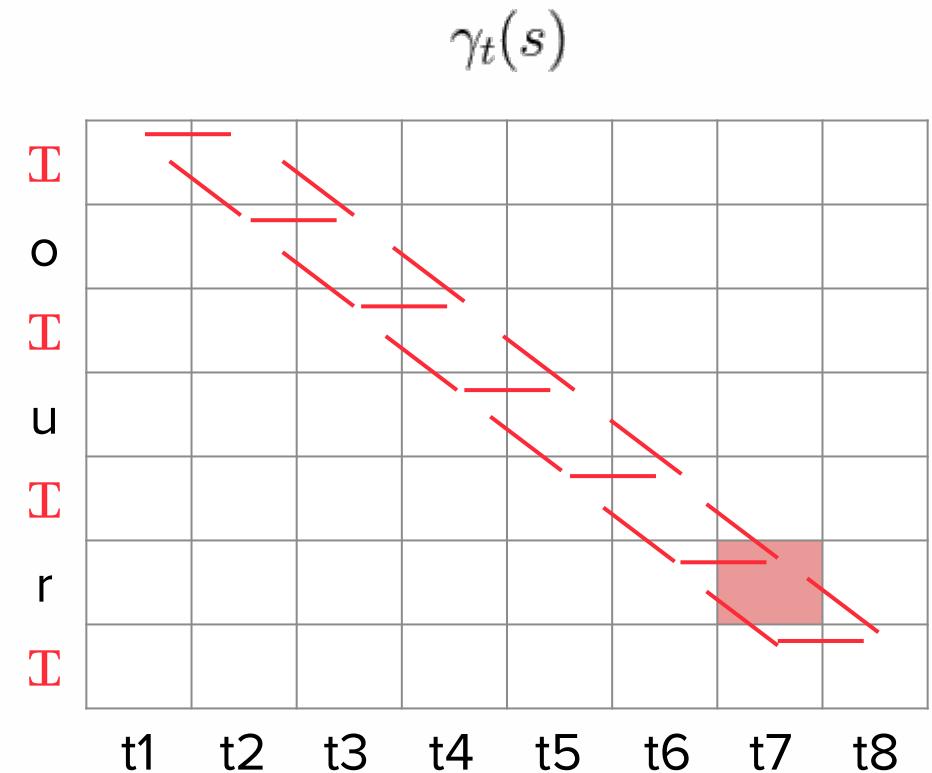
Градиент СТС

$\gamma_t(s)$ - вероятность всех путей, проходящих (t, s)

$$L'_{p_7(r)} = -\frac{1}{P(label|pred)} \sum_{s:c(s)=r} \frac{\gamma_7(s)}{p_7(r)}$$

T	0.9	0.2	0.8	0.1	0.3	0.7	0.3	0.6
O	0.1	0.8	0.1	0.1	0.0	0.0	0.0	0.0
R	0.0	0.0	0.0	0.0	0.1	0.2	0.7	0.4
U	0.0	0.0	0.1	0.8	0.6	0.1	0.0	0.0

pred



Вопросы

Оценка OCR

Оценка OCR

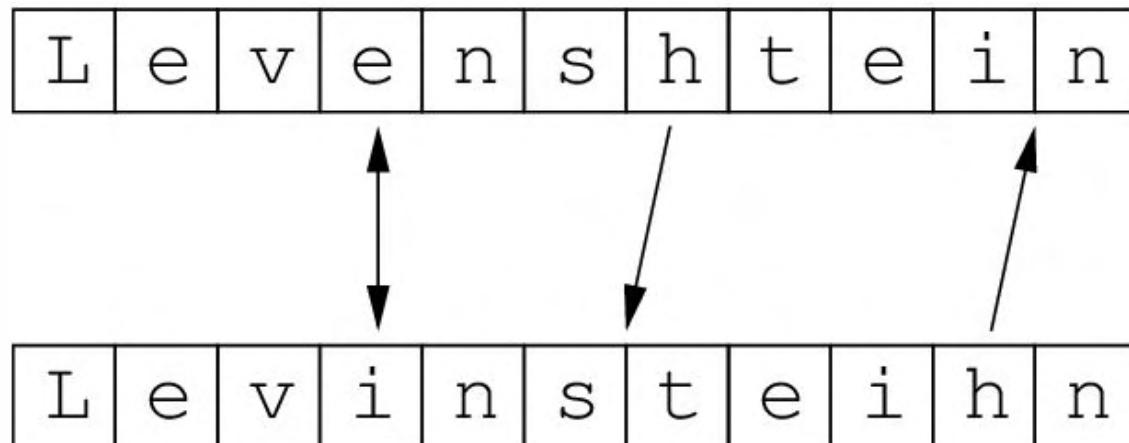
Редакционное расстояние: замена, удаление, вставка

[Про вычисление см. Википедию](#)



Метрики:

- Доля неверно распознанных символов (CER)
- Доля неверно распознанных слов (WER)



$$\begin{aligned} D &= 3 \\ \text{CER} &= 3 / 11 \end{aligned}$$

Задание

Тест про редакционное расстояние

❖ Ссылка в чате

☒ 3 минуты



L	e	v	e	n	s	h	t	e	i	n
---	---	---	---	---	---	---	---	---	---	---

L	e	v	i	n	s	t	e	i	h	n
---	---	---	---	---	---	---	---	---	---	---



Заключение



Feedback



Пишите в чат или говорите голосом

- * Что запомнилось?

- * Что трудно и вызывает вопросы?

- * Про что хотелось бы услышать ещё?



Цели

- 1 Получить теорию для семинара
- 2 Научиться вычислять CTC loss
- 3 Вспомнить архитектуры сетей для OCR



Спасибо!

Семинар после перерыва