

Применение генеративных текстовых моделей для решения задач обработки естественного языка

Васильев Руслан

ММП ВМК МГУ, 2021

Содержание

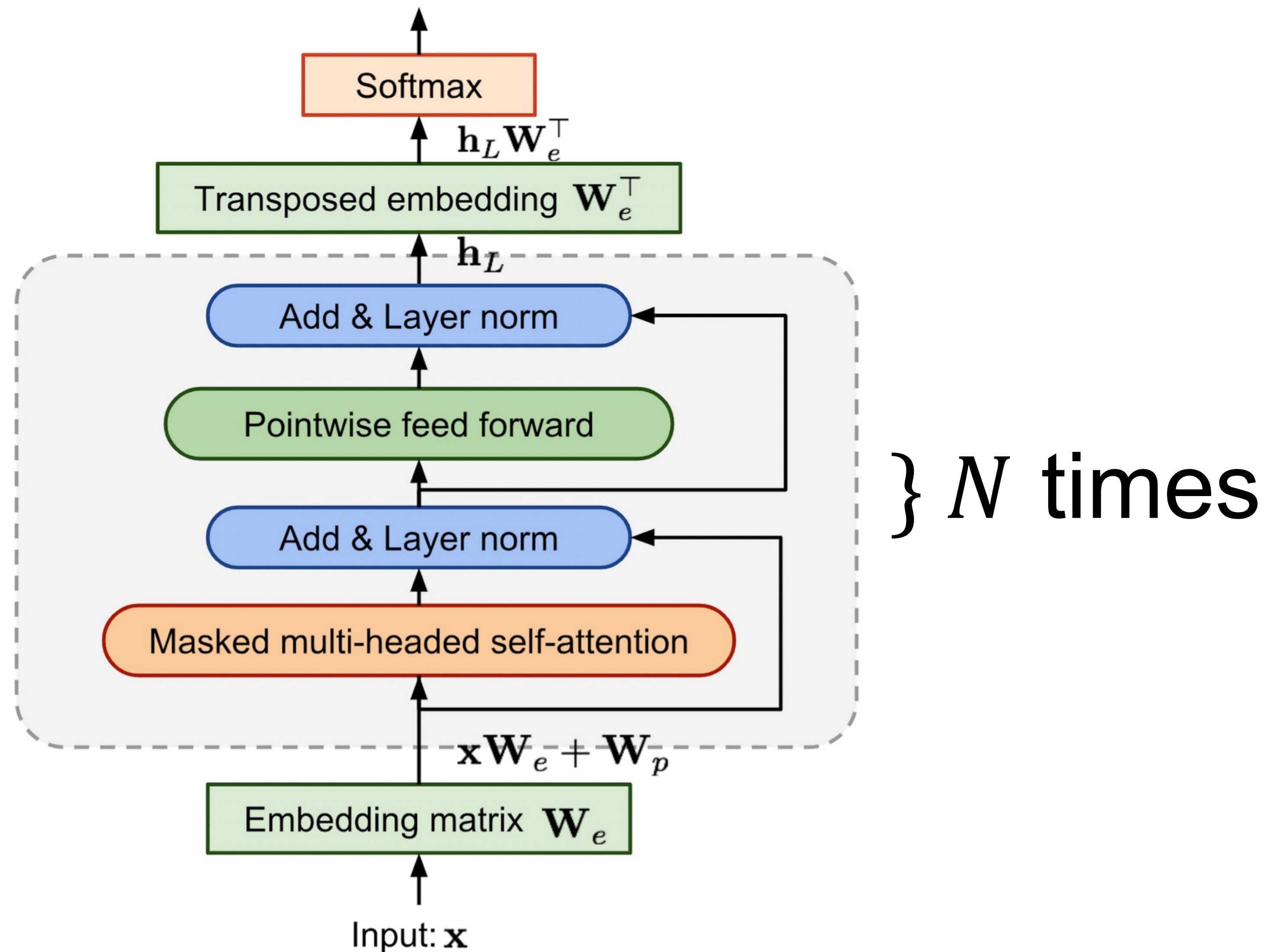
1. Задачи
2. Метод решения
3. Эксперименты
4. Результаты

Языковая модель

$$p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i \mid x_{<i})$$

$$L_{LM} = - \sum_{i=1}^n \log p(x_i \mid x_{<i})$$

Семейство YaLM



Russian SuperGLUE

	Task type	Metrics	Train / Val / Test size
LiDiRus	NLI & diagnostics	Matthews Corr	0 / 0 / 1104
RCB	NLI	Avg. F1 / Accuracy	438 / 220 / 438
PARus	Common Sense	Accuracy	400 / 100 / 500
MuSeRC	Machine Reading	F1a / EM	500 / 100 / 322
TERRa	NLI	Accuracy	2616 / 307 / 3198
RUSSE	Common Sense	Accuracy	19845 / 8508 / 18892
RWSD	Reasoning	Accuracy	606 / 204 / 154
DaNetQA	World Knowledge	Accuracy	1749 / 821 / 805
RuCoS	Machine Reading	F1 / EM	72193 / 7577 / 7257

Пример: MuSeRC

(1) Зайчик взял с собой несколько пирожков и отправился в путь. (2) Долго-долго шел он по лесу. (3) Ничто не могло сломить его, потому что зайчик хорошо кушал и был сильным. (4) Когда он чувствовал, что у него заканчивались силы, он съедал пирожок, и силы сразу возвращались. (5) Когда все пирожки были съедены, зайчонок как раз дошёл до своего дома. (6) Открыв дверь, он увидел свою маму, которая сидела у окна и горько плакала. (7) - Мамочка! – позвал её зайчонок, - я вернулся! (8) Мама обернулась и долго смотрела на него, потому что не узнала сразу. (9) Десять дней назад зайчонок был худеньким и слабеньким, как пёрышко. (10) А сейчас перед зайчихой стоял сильный и крепкий заяц, и маме даже показалось, что он стал выше ростом. (11) - Зайчонок! (12) Неужели это ты! – обрадовалась мама. (13) Она подбежала к своему сыночку и стала обнимать и целовать его. (14) - Давай покушаем? – предложил зайчонок. (15) С тех пор зайчик очень хорошо кушает. (16) А ещё он любит помогать маме в готовке.

Почему мама
не узнала зайчонка?

Зайчонок подрос и
поправился.

Он был грязным.

Он изменился.

Он был слишком худой.

MuSeRC: few/zeroshot

Прочитайте отрывок и ответьте на вопрос.

<text>

Вопрос:

<question>

Уточняющий
вопрос:

<answer>?

Да

Нет

P-tuning

$$p\left(\vec{h}, \vec{x}, \overrightarrow{h'}, \vec{y}\right)$$

$$p\left(\vec{y} \mid \vec{h}, \vec{x}, \overrightarrow{h'}\right) \rightarrow \max_{\vec{h}, \overrightarrow{h'}}$$

$$L = - \sum_{i=1}^n \log p(y_i | y_{<i}, \vec{h}, \vec{x}, \overrightarrow{h'})$$

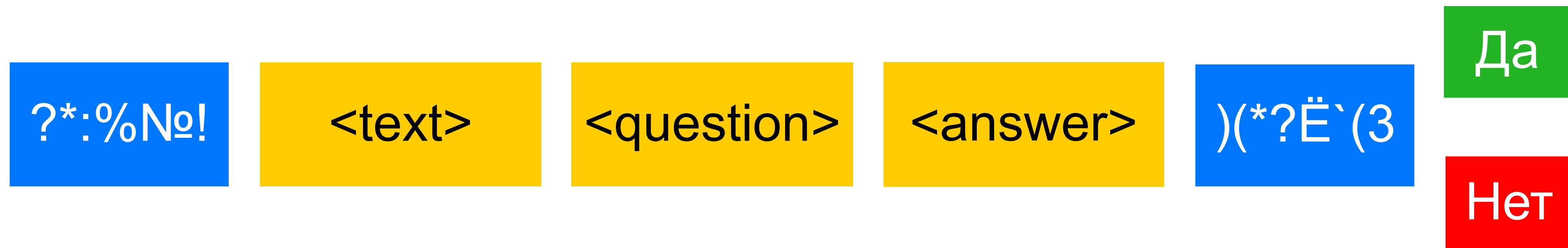
~@)*#&`+`

<prefix>

?*:%№!

<suffix>

MuSeRC: p-tuning



PARus

Дом отбрасывает
тень на траву.

причина

Солнце уже
поднялось.

Трава уже
подстрижена.

Культуристка
подняла вес.

следствие

Её мышцы
устали.

Тренажёрный
зал закрылся.

PARus: p-tuning 1



PARus: p-tuning 2

причина

?*:%№!

<premise>

)(*?Ё̀(3

<choice 1>

<choice 2>

следствие

?*:%№!

<premise>

)(*?Ё̀(3

<choice 1>

<choice 2>

PARus: p-tuning 3

причина

?*:%№!

<choice 1>

<choice 2>

)(*?Ё̀(3

<premise>

следствие

?*:%№!

<choice 1>

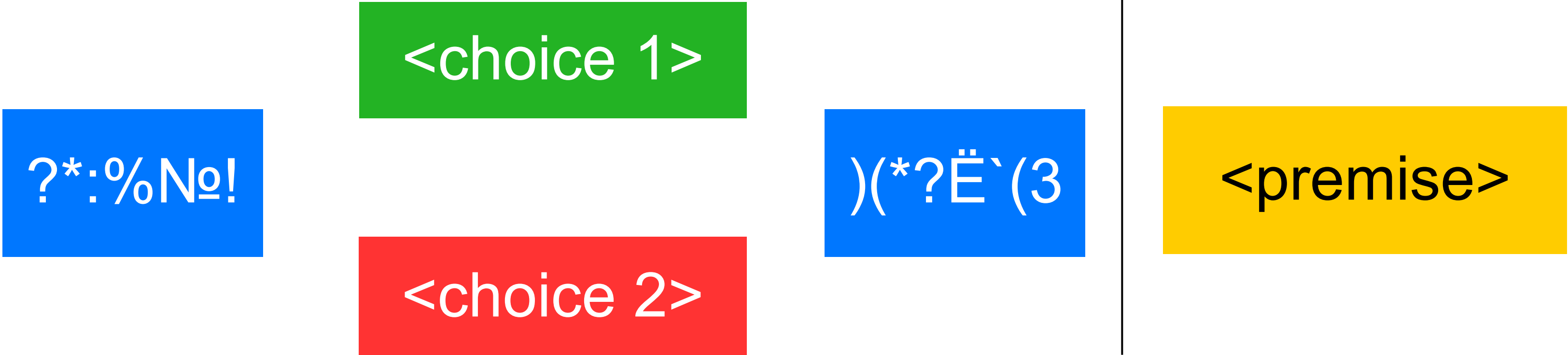
<choice 2>

)(*?Ё̀(3

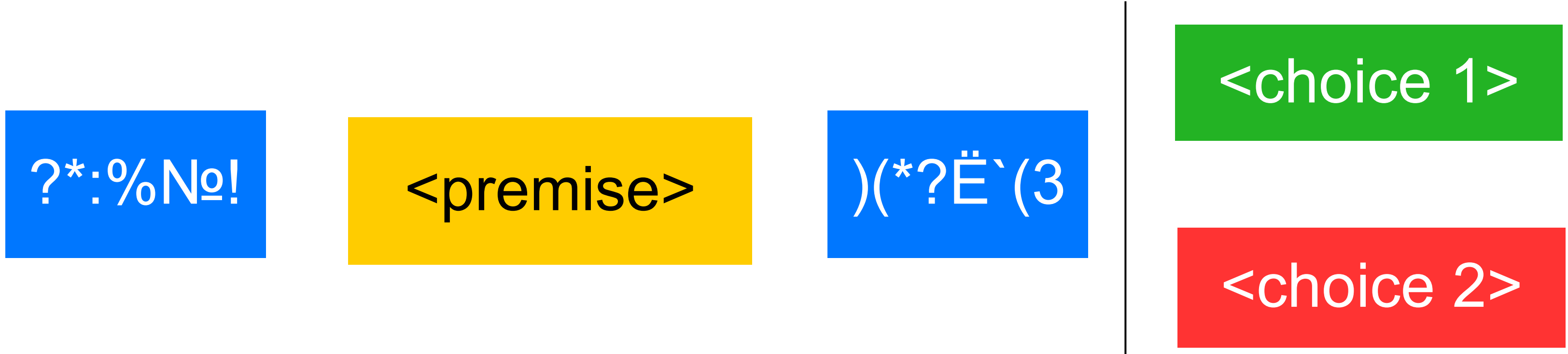
<premise>

PARus: p-tuning 4

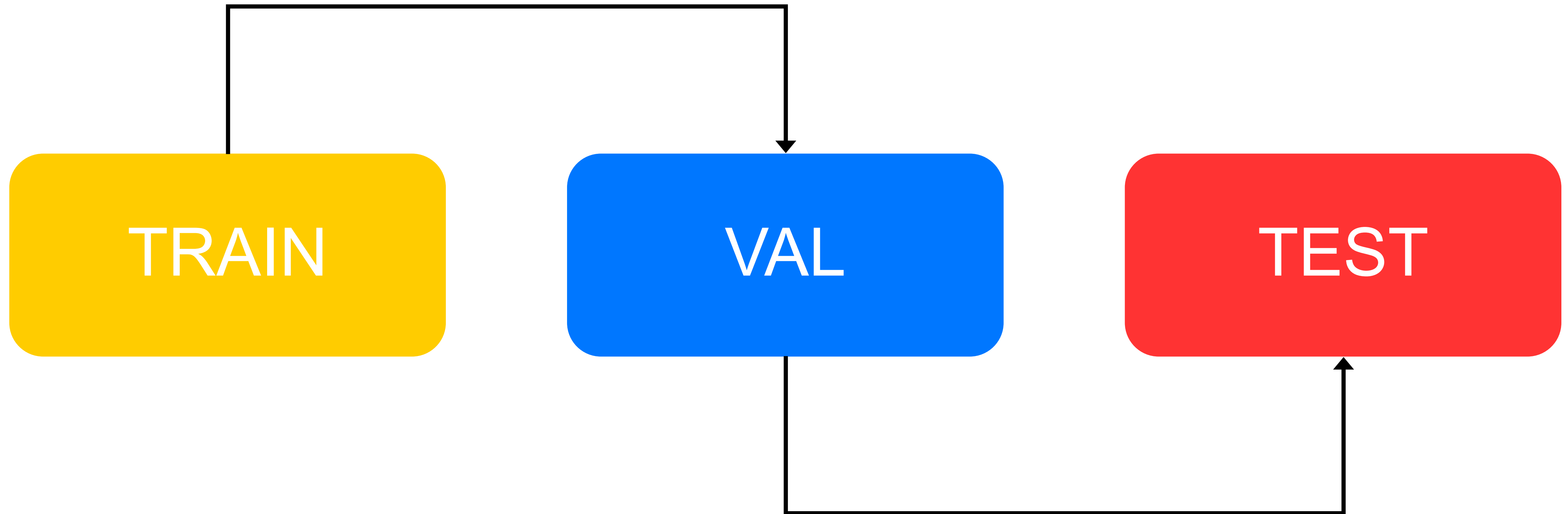
причина

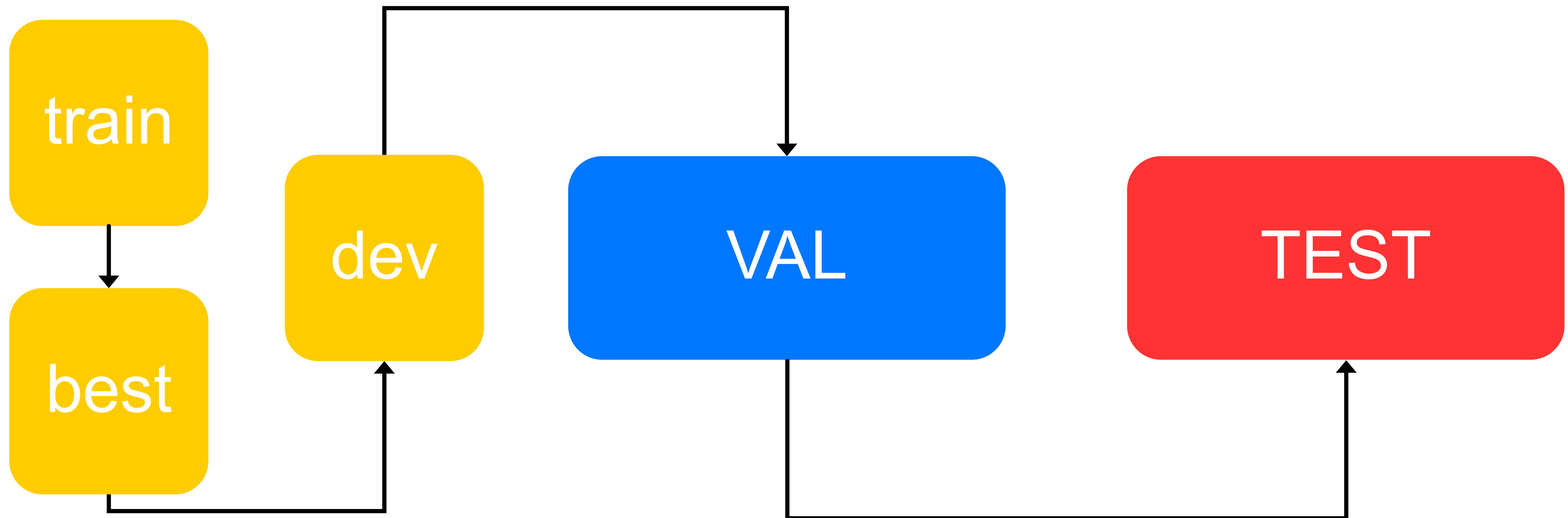


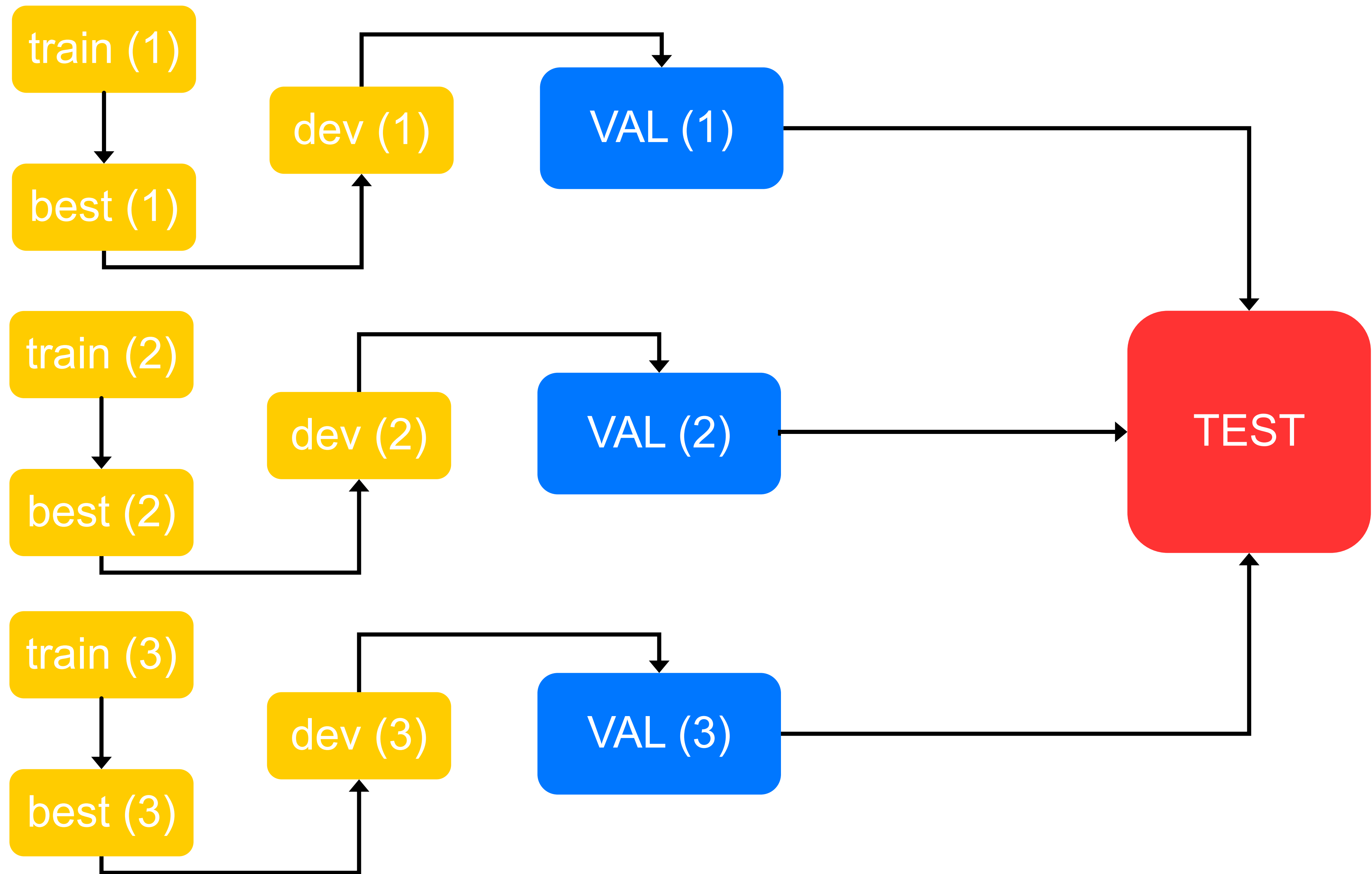
следствие



Валидация





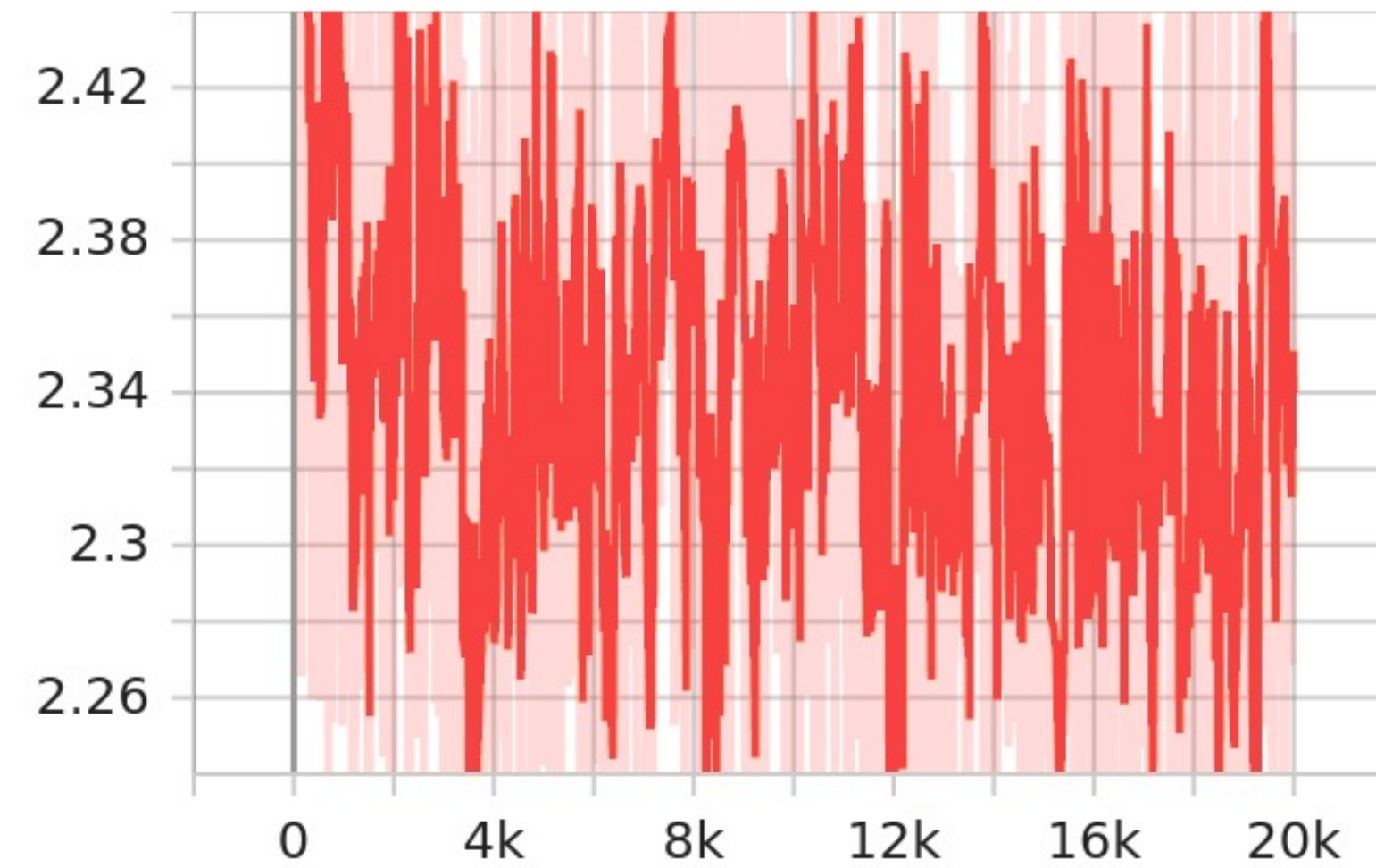


Обучение

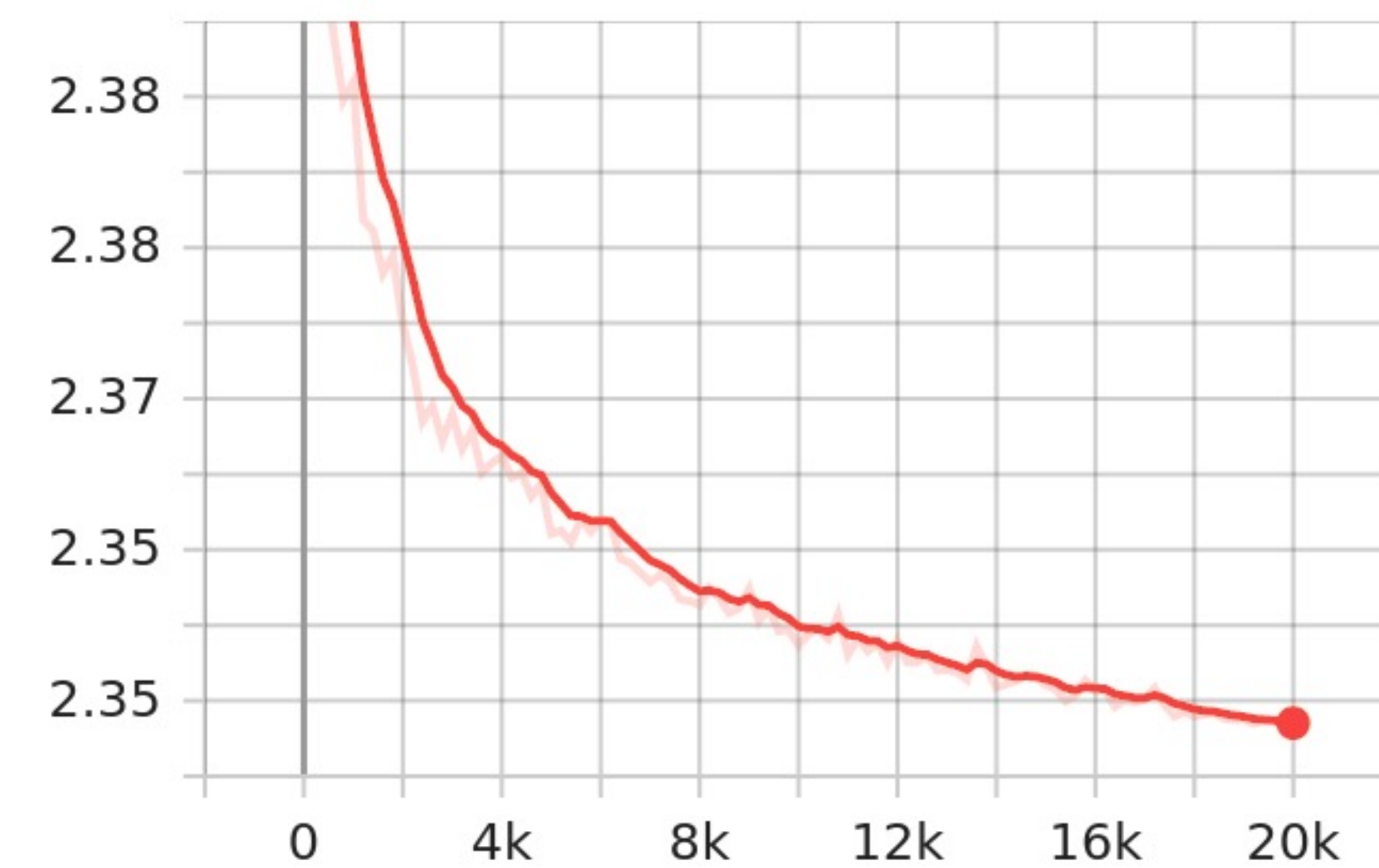
high-res

large suffix

train loss

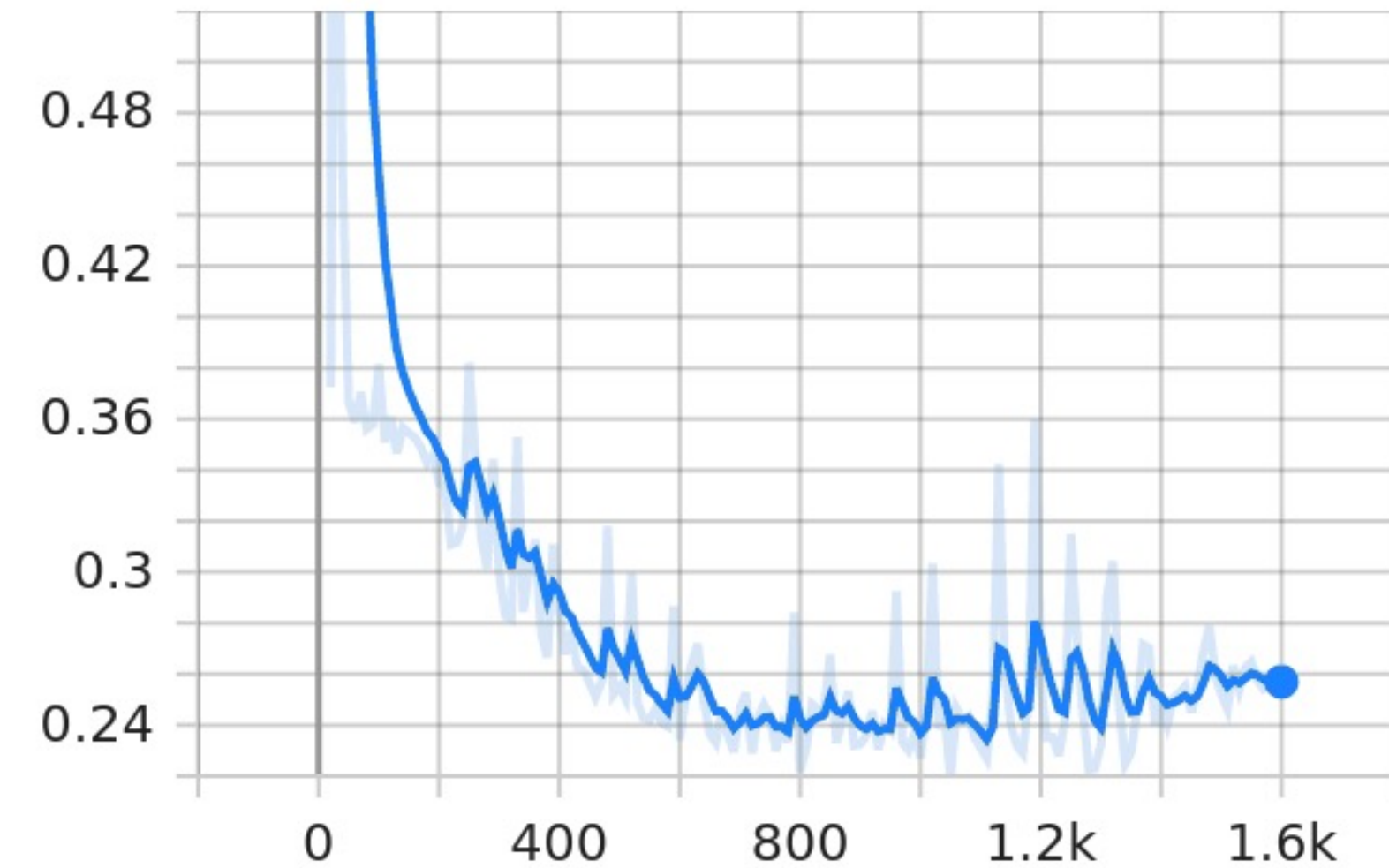
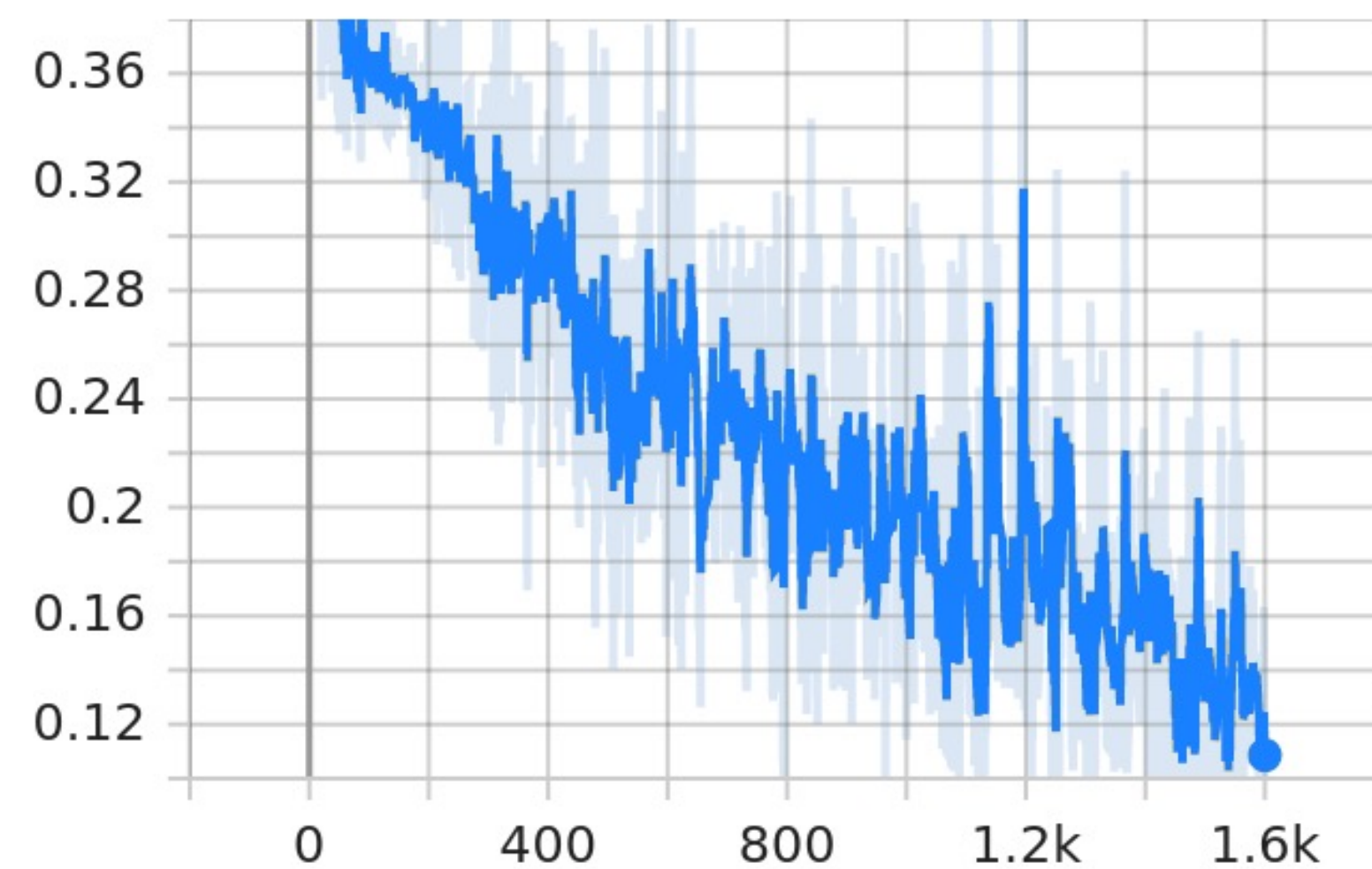


val loss



low-res

small suffix



Гиперпараметры

lr

min lr

train iters

lr decay iters

lr decay style

warmup

weight decay

ptune dropout

hidden dropout

attention dropout

init embedding

block size

use mlp

use lstm

Инициализация

1. Аналогично инициализации эмбеддингов
2. Случайными токенами из словаря
3. Нормальным распределением с дисперсией обученных токенов
4. Токенами, связанными с задачей
5. Осмысленной подводкой

Результаты (test)

	score	LiDiRus	RCB	PARus	MuSeRC	TERRa	RUSSE	RWSD	DaNetQA	RuCoS
1. Human Benchmark	0.811	0.626	0.680 / 0.702	0.982	0.806 / 0.420	0.92	0.805	0.84	0.915	0.930 / 0.890
2. GT v2	0.755	0.515	0.384 / 0.534	0.906	0.936 / 0.804	0.877	0.687	0.643	0.911	0.920 / 0.924
3. YaLM (p-tuning)	0.711	0.364	0.357 / 0.479	0.834	0.892 / 0.707	0.841	0.710	0.669	0.85	0.920 / 0.916
4. ruT5 (fine-tuning)	0.686	0.32	0.45 / 0.532	0.764	0.855 / 0.608	0.775	0.775	0.669	0.79	0.860 / 0.859
5. ruRoberta (fine-tuning)	0.684	0.343	0.357 / 0.518	0.722	0.861 / 0.630	0.801	0.748	0.669	0.82	0.870 / 0.867
...										
10. YaLM (few-shot)	0.577	0.124	0.408 / 0.447	0.766	0.673 / 0.364	0.605	0.587	0.669	0.637	0.860 / 0.859
...										
...										
23. Majority class	0.374	0.0	0.217 / 0.484	0.498	0.0 / 0.0	0.513	0.587	0.669	0.503	0.25 / 0.247

Итоги

1. Сабмит на Russian SuperGLUE
2. Как (не) надо учить p-tune вектора
3. Как решать любые* языковые задачи
4. Оптимальная параметризация и гиперпараметры