

# Методы предобработки текстовых данных для ускорения обучения языковых моделей

Сурков Максим Константинович

Научный руководитель: Ямщиков Иван Павлович

Санкт-Петербургская школа физико-математических и компьютерных наук  
НИУ ВШЭ СПб

17 марта 2021 г.

# Обработка естественного языка в реальной жизни

- социальные сети
- электронная почта
- службы доставки
- голосовые помощники
- переводчики
- чат боты



- ❶ классификация последовательностей
  - спам
  - грубая речь<sup>1</sup>
- ❷ генерация выходной последовательности из исходной
  - машинный перевод
  - ответы на вопросы
- ❸ выделение информации из последовательностей
  - выделение именованных сущностей<sup>2</sup>

---

<sup>1</sup>G. H. Paetzold et al., SemEval'19 Task 5: Hate Speech Identification with RNN.

<sup>2</sup>Vikas Yadav et al., SemEval'19 Task 12: Deep-Affix Named Entity Recognition of Geolocation Entities. ACL'19

# Современные методы решения задач обработки естественного языка

- ❶ Механизм внимания<sup>1</sup>
- ❷ **BERT** (Google)<sup>2</sup>
- ❸ GPT-3 (OpenAI)<sup>3</sup>

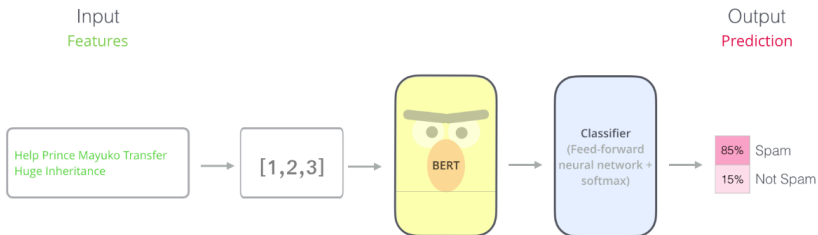
---

<sup>1</sup>Ashish Vaswani et al., Attention Is All You Need, 2017

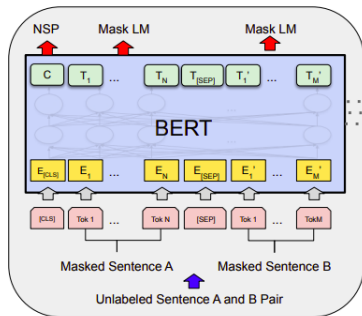
<sup>2</sup>Jacob Devlin et al., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2019

<sup>3</sup>Tom B. Brown et al., Language Models are Few-Shot Learners, 2020

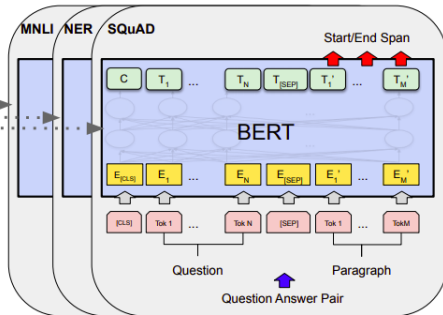
# BERT. Использование



# BERT. Обучение



Pre-training



Fine-Tuning

# BERT. Требуемые ресурсы

- количество параметров: 110M – 340M
- время на предобучение: от 2-4 дней до 1-2 недель<sup>1</sup>
  - мировой рекорд: 47 минут на **1472** V100 GPU<sup>2</sup>
- время на дообучение: 1-2 дня
- размеры данных:

Датасет	Размер
Wikipedia	3-600M
HND	600k-2M
s140	1.6M
IWSLT	200-230k
QQP	364k
MNLI	393k

<sup>1</sup>При использовании 1x-4x GPU Nvidia Tesla V100 32Gb

<sup>2</sup><https://developer.nvidia.com/blog/training-bert-with-gpus>

# BERT. Существующие методы оптимизации

- квантизация<sup>1</sup>
- дистилляция<sup>2</sup>
- прунинг<sup>3</sup>

---

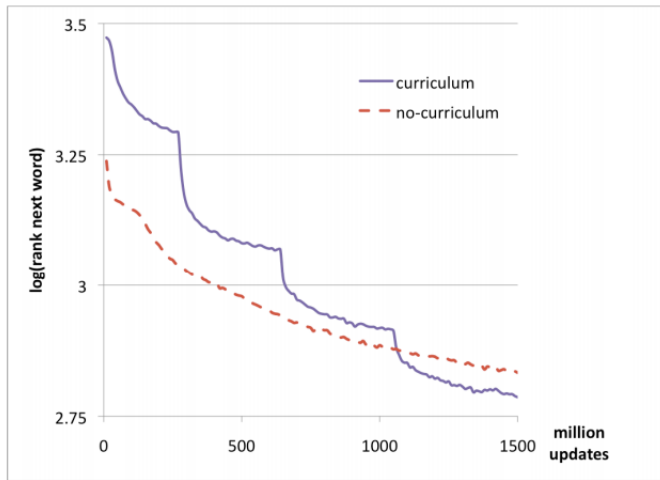
<sup>1</sup>Sheng Shen et al., Q-BERT: Hessian Based Ultra Low Precision Quantization of BERT, 2019

<sup>2</sup>Victor Sanh et al., DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter, 2020

<sup>3</sup>Hassan Sajjad et al., Poor Man's BERT: Smaller and Faster Transformer Models, 2020

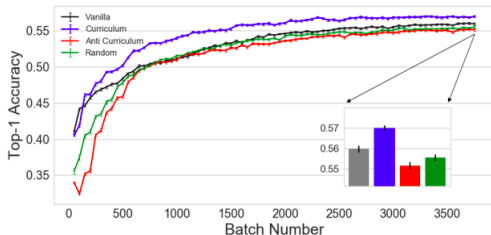


# Обучение с расписанием. Начало



Y. Bengio et al., Curriculum learning, 2009

- компьютерное зрение<sup>1</sup>



- обучение с подкреплением<sup>2</sup>

- глубокое обучение<sup>3</sup>

<sup>1</sup>Guy Hach Cohen, Daphna Weinshall, On The Power of Curriculum Learning in Training Deep Networks, 2019

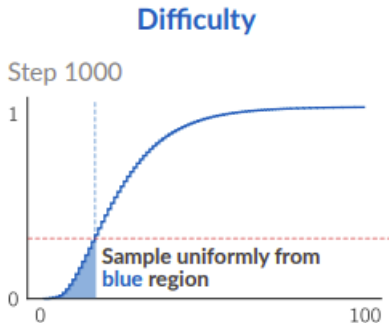
<sup>2</sup>Sanmit Narvekar et al., Curriculum Learning for Reinforcement Learning Domains: A Framework and Survey, 2020

<sup>3</sup>Mermer et al., Scalable Curriculum Learning for Artificial Neural Networks, 2017

# Обучение с расписанием в обработке языка

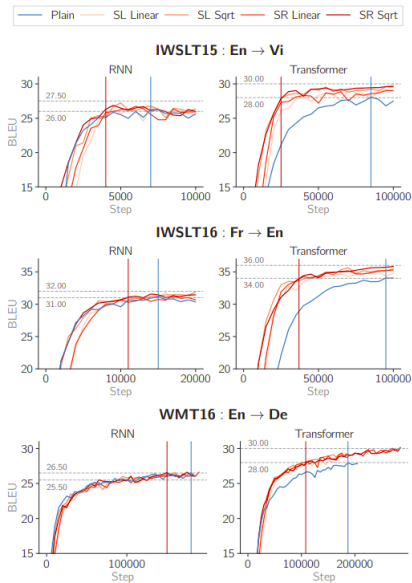
- Задача: машинный перевод
- Модель: BERT, LSTM
- Датасеты: IWSLT'15, IWSLT'16, WMT'16
- Алгоритм:

- 1 сортируем тексты по сложности (длина, логарифм вероятности правдоподобия)
- 2 в течение  $T$  шагов (рассмотрим шаг  $t$ )
  - считаем  $c(t) \in [0, 1]$
  - строим батч из  $c(t)$  **первых** текстов корпуса
  - шаг обучения



Е. А. Platanios et al., Competence-based Curriculum Learning for Neural Machine Translation, ACL'19

# Обучение с расписанием в обработке языка



---

Benfeng Xu et al., Curriculum Learning for Natural Language Understanding,  
ACL'20

# Обучение с расписанием в обработке языка.

## Направления для исследований