Вся эта BERTология

Татьяна Шаврина
AGI NLP, Sberdevices

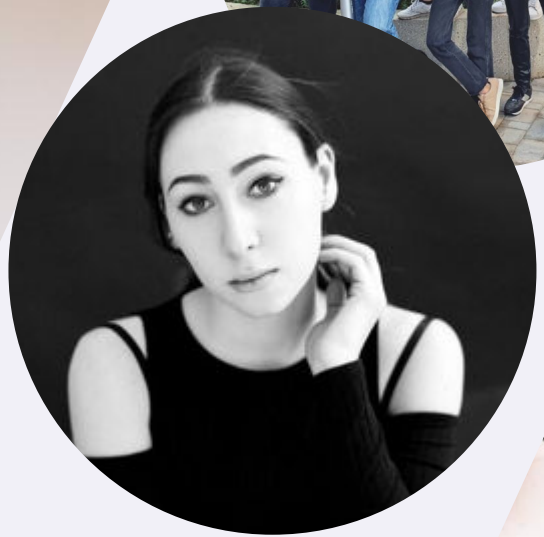
HELLO!

I am Tatiana Shavrina

I am here because I am an NLP enthusiast

You can find me at @rybolos

AGI NLP team



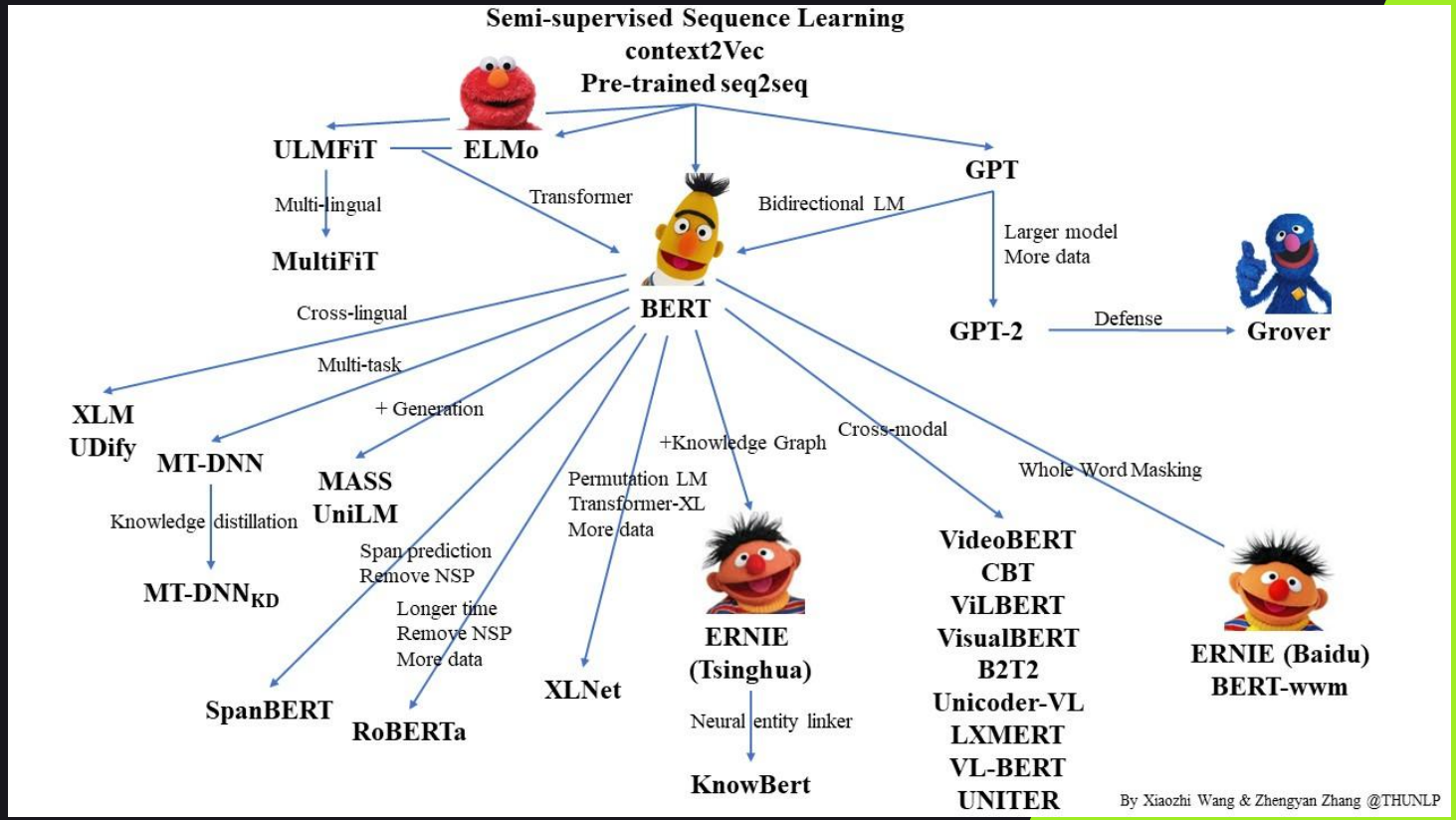
Sberdevices

Model Zoo

BERT, GPT-3...

pretrained models

universal abilities to recreate human skills



2017 - 2020 - Human performance achieved

Microsoft researchers achieve new conversational speech recognition milestone

Published August 20, 2017

By Xuedong Huang, Technical Fellow and Chief Technology Officer Azure C
Services

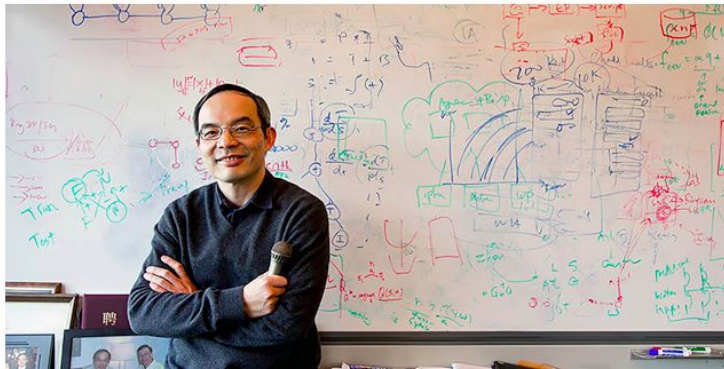
Microsoft creates AI that can read a document and answer qu
about it as well as a person

January 15, 2018 | Allison Linn

NEWS

Microsoft reaches a historic milestone, usin AI to match human performance in translating news from Chinese to English

TEAM TTR - 16/03/2018



Transformers are all we need

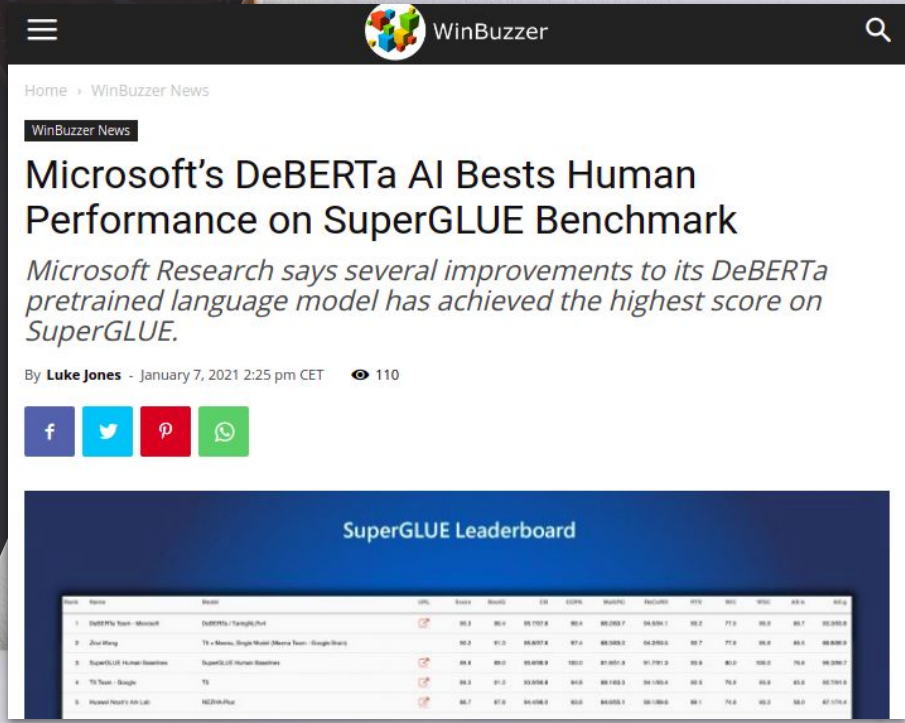
SOTA results with transformers:

- Open-Domain Question Answering
- Sentiment Classification
- Machine Translation
- Text Generation
- Named Entity Recognition
- Reading Comprehension
- General Language Understanding
- and much more...

Transformers are all we need

SOTA results with transformers:

- Open-Domain Question Answering
- Sentiment Classification
- Machine Translation
- Text Generation
- Named Entity Recognition
- Reading Comprehension
- General Language Understanding
- and much more...



First Russian Transformer Models

RuBERT, Russian, cased, 12-layer, 768-hidden, 12-heads, 180M parameters

SlavicBERT, Slavic (bg, cs, pl, ru), cased, 12-layer, 768-hidden, 12-heads, 180M parameters

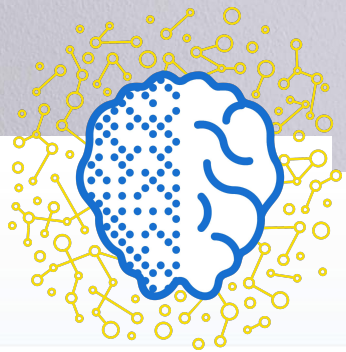
Conversational BERT, English, cased, 12-layer, 768-hidden, 12-heads, 110M parameters

Conversational RuBERT, Russian, cased, 12-layer, 768-hidden, 12-heads, 180M parameters

Sentence Multilingual BERT, 101 languages, cased, 12-layer, 768-hidden, 12-heads, 180M parameters

Sentence RuBERT, Russian, cased, 12-layer, 768-hidden, 12-heads, 180M parameters

First Russian Transformer Models



Hugging Face

Search models, datasets, users...

DeepPavlov / **rubert-base-cased**

PyTorch | ru | arxiv:1905.07213 | bert

Model card | Files and versions

rubert-base-cased

RuBERT (Russian, cased, 12-layer, 768-hidden, 12-heads, 180M parameters) was tra

part of Wikipedia and news data. We used this training data to build a vocabulary of Russian subtokens

and took a multilingual version of BERT-base as an initialization for RuBERT[1].

Downloads last month

17,289

Hosted inference API

Unable to determine this model's pipeline type

NTI AI Hackaton

Искусственный интеллект

АКАДЕМИЯ

ВКЛАД В БУДУЩЕЕ

Олимпиада НТИ

01. Задача

Реши задачи по NLP лучше других и докажи, что достоин забрать главный приз

02. Чат

Общаемся, обсуждаем новости, задаем вопросы организаторам в Телеграмм чате

03. Рейтинг

rubert_conv_dp_notlower	rubert_sen_dp_lower
германии	германии
хамас	рф
ташкента	ташкента
сми	heckler & koch
франк-вальтер штайнмайер	франк-вальтер штайнмайер

```
о один из наших методов ансамблирования
S. уважаемые программисты, не делайте сразу фейспа
гласен, это сложно назвать нормальным ансамблем
 оно повышает точность и использует несколько моде
an = []
```

хафтар	халифа хафтар
россии	хафтара
мазиной	мазиной

Загрузка обученных нами моделей

```
[ ]: # with open('/content/drive/MyDrive/New_models/bert_f
     #     bert_xquad_notlower = pickle.load(f)
     # with open('/content/drive/MyDrive/New_models/distil
     #     distilbert_notlower = pickle.load(f)
     with open('/content/drive/MyDrive/dpmlbert_cased_lowe
         bert_dp_lower = pickle.load(f)
     with open('/content/drive/MyDrive/dprubertconv_cased_
         rubert_conv_dp_lower = pickle.load(f)
     with open('/content/drive/MyDrive/dprubertconv_cased_
         rubert_conv_dp_notlower = pickle.load(f)
     with open('/content/drive/MyDrive/model_rubert_low.pk
         rubert_lower = pickle.load(f)
     with open('/content/drive/MyDrive/model_rubert_senten
         rubert_sen_lower = pickle.load(f)
     with open('/content/drive/MyDrive/model_rubert_no.pkl
         rubert_sen_dp_lower = pickle.load(f)

     with open('/content/drive/MyDrive/finalized_model_ber
         bert_fin = pickle.load(f)
```

А дальше начинается сущий ад и куча методов ансамблирования предиктов бертов, которые мы придумали

```
[ ]: # получение предиктов каждого берта
```

NTI AI Hackaton

Искусственный интеллект

АКАДЕМИЯ

ВКЛАД В БУДУЩЕЕ

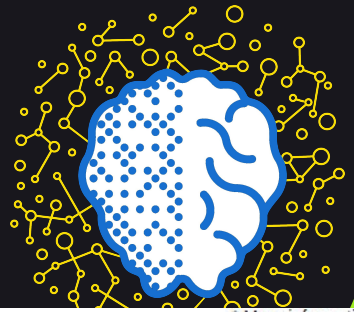
Олимпиада НТИ

03. Рейтинг

Обр. хакатон | Отборочный | Финал

Место	Имя	Результат
1	Avengers Ensemble	0.9313
2	Братва рвется в топ	0.8847
3	Спутник-V	0.8753
4	Почему Берт выдаёт единички	0.8693
5	{team_name}	0.8693
6	RuGpt3	0.864
7	Arima	0.8573
8	The AI Gang	0.8533
9	Ninja Turtles	0.8447
10	NTI: Become chelovek	0.8333

First Models on Russian SuperGLUE



* More information about speed sc

Rank	Name	Team	Link	Score	LiDiRus	RCB	PARus	MuSeRC	TERRa	RUSSE	RWSD	DaNetQ
1	HUMAN BENCHMARK	AGI NLP	i	**0.811**	0.626	0.68 / 0.702	0.982	0.806 / 0.42	0.92	0.805	0.84	0.915
2	RuGPT3XL few-shot	sberdevices	i	**0.535**	0.096	0.302 / 0.418	0.676	0.74 / 0.546	0.573	0.565	0.649	0.59
3	MT5 Large	AGI NLP	i	**0.528**	0.061	0.366 / 0.454	0.504	0.844 / 0.543	0.561	0.633	0.669	0.657
4	RuBERT plain	DeepPavlov	i	**0.521**	0.191	0.367 / 0.463	0.574	0.711 / 0.324	0.642	0.726	0.669	0.639
5	RuGPT3Large	sberdevices	i	**0.505**	0.231	0.417 / 0.484	0.584	0.729 / 0.333	0.654	0.647	0.636	0.604
6	RuBERT conversational	DeepPavlov	i	**0.5**	0.178	0.452 / 0.484	0.508	0.687 / 0.278	0.64	0.729	0.669	0.606
7	Multilingual Bert	DeepPavlov	i	**0.495**	0.189	0.367 / 0.445	0.528	0.639 / 0.239	0.617	0.69	0.669	0.624
8	heuristic majority	ling_ling	i	**0.468**	0.147	0.4 / 0.438	0.478	0.671 / 0.237	0.549	0.595	0.669	0.642
9	RuGPT3Medium	sberdevices	i	**0.468**	0.01	0.372 / 0.461	0.598	0.706 / 0.308	0.505	0.642	0.669	0.634
10	RuGPT3Small	sberdevices	i	**0.438**	-0.013	0.356 / 0.473	0.562	0.653 / 0.221	0.488	0.57	0.669	0.61
11	Baseline TF-IDF1.1	AGI NLP	i	**0.434**	0.06	0.301 / 0.441	0.486	0.587 / 0.242	0.471	0.57	0.662	0.621

BERTology

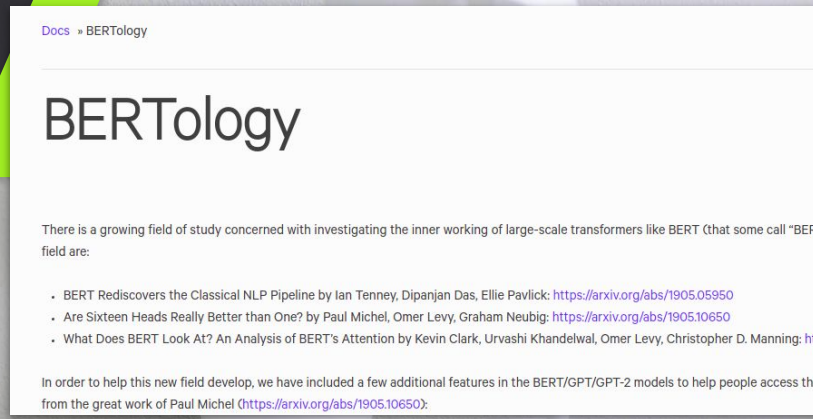

BERTology
Ways to look into
the black void

What does Bertology do?

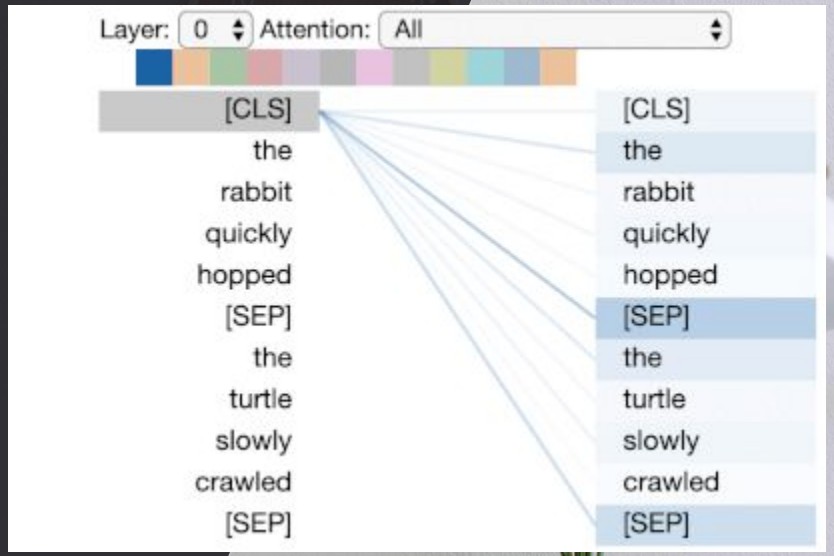
- accessing all the hidden-states of BERT/GPT/GPT-2,
- accessing all the attention weights for each head of BERT/GPT/GPT-2, 3...
- retrieving heads output values and gradients to be able to compute head importance score
- probing! evaluate layer representations

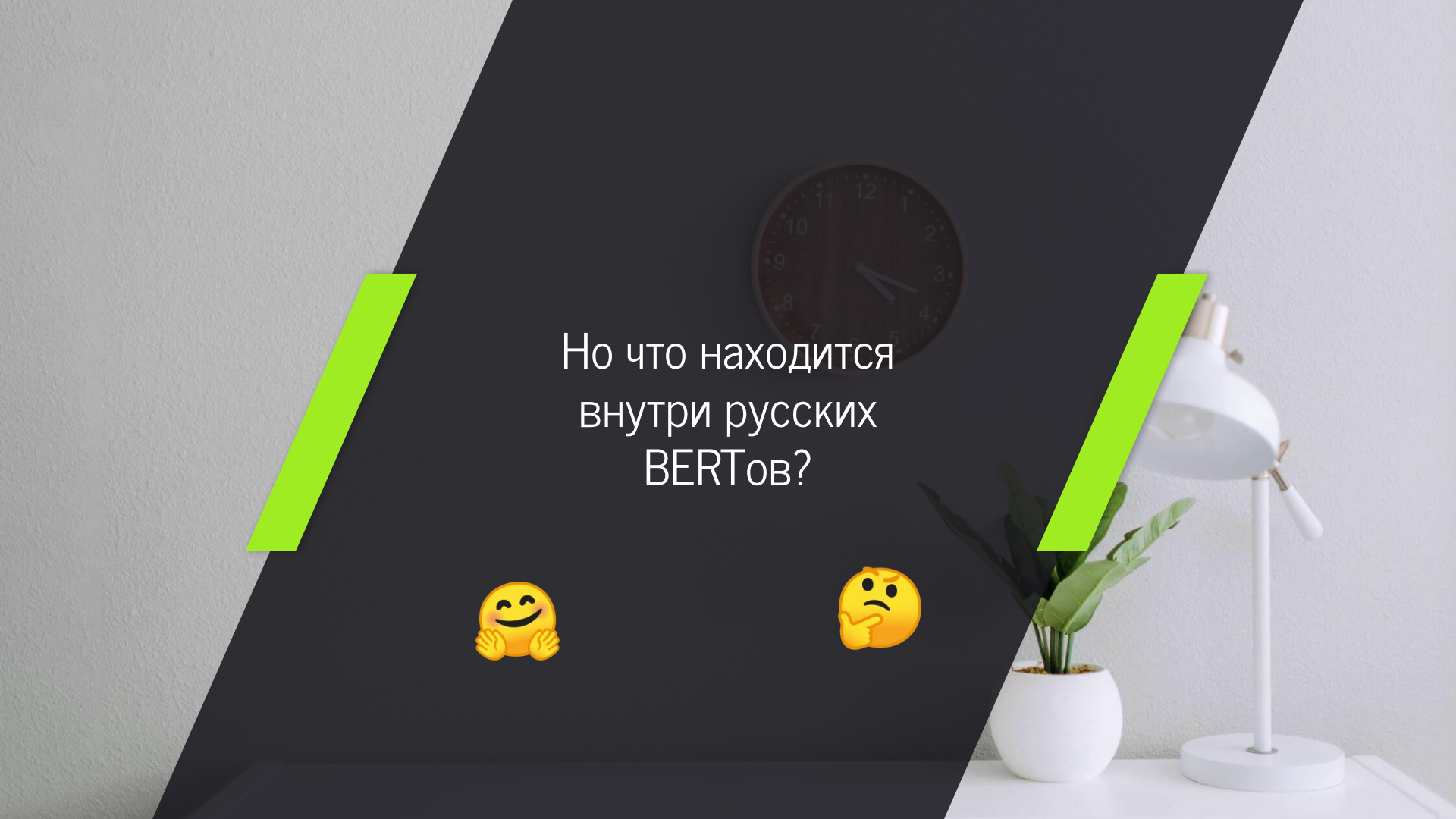
What does Bertology do?

- accessing all the hidden-states of BERT/GPT/GPT-2,
- accessing all the attention weights for each head of BERT/GPT/GPT-2, 3...
- retrieving heads output values and gradients to be able to compute head importance score
- probing! evaluate layer representations



What does Bertology do?




Но что находится
внутри русских
BERTов?

RuSentEval framework

probing Russian models

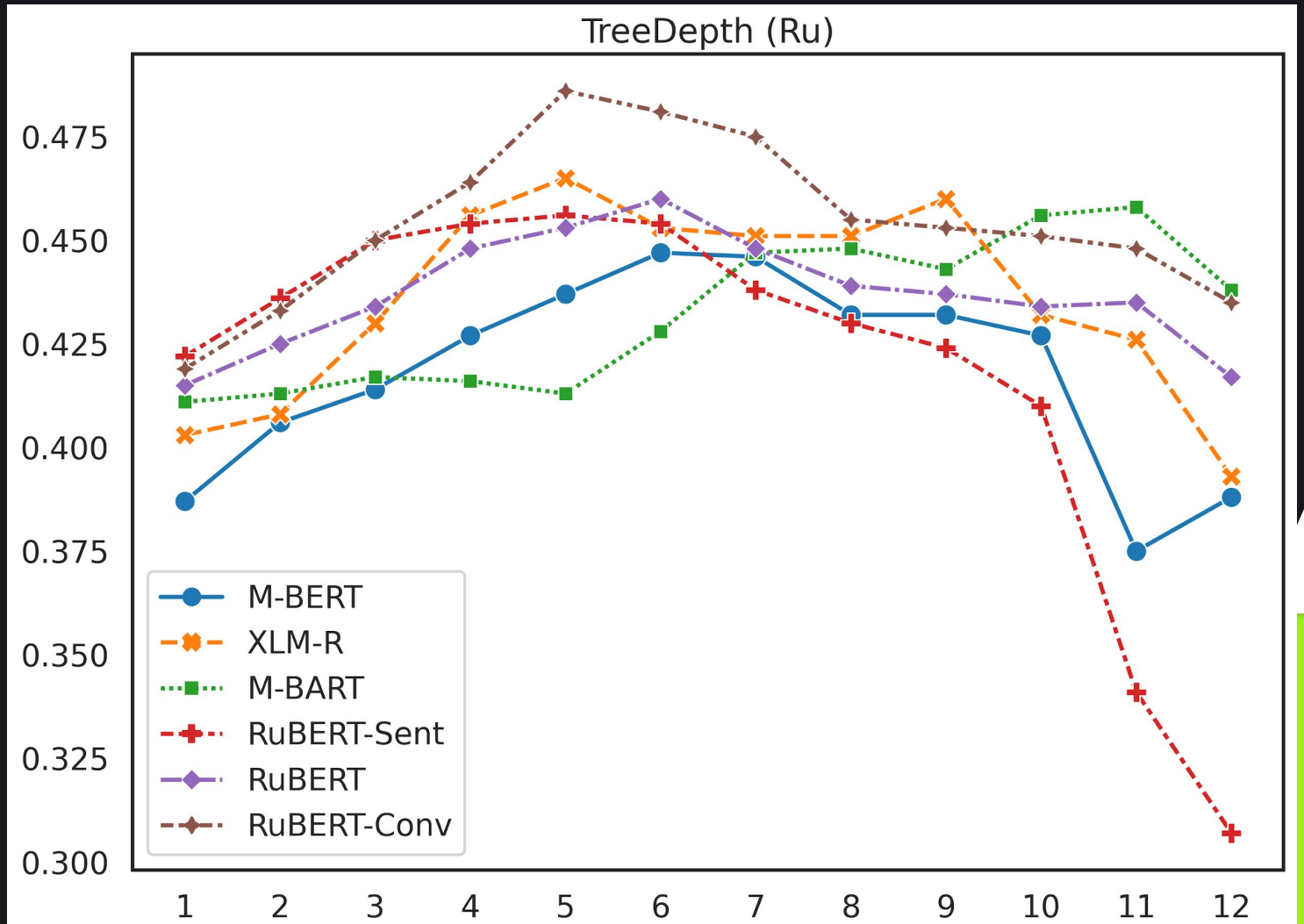
RuSentEval - First Russian Probing

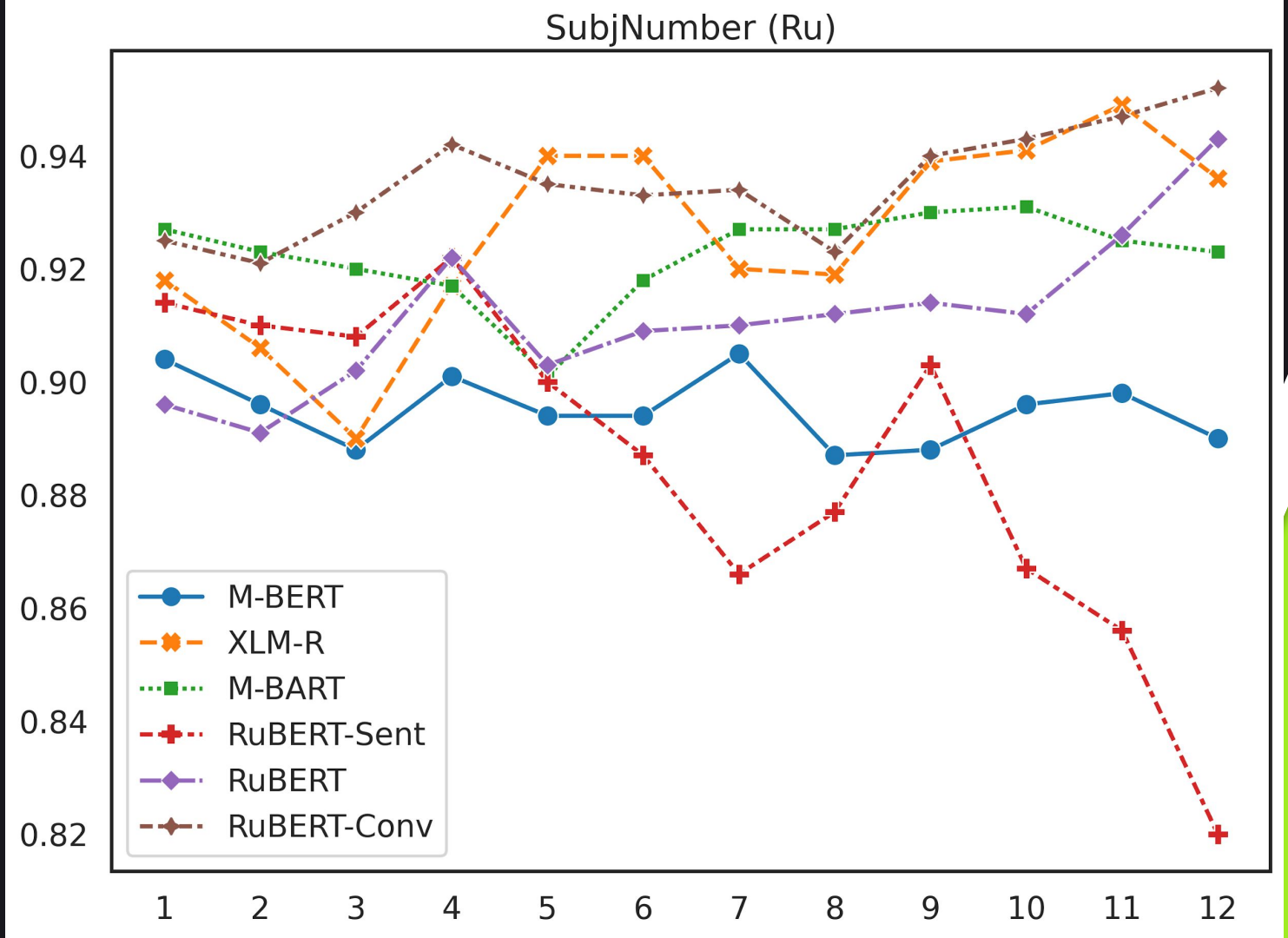
BERT-like models - source of embeddings:
word embeddings, sentence embeddings

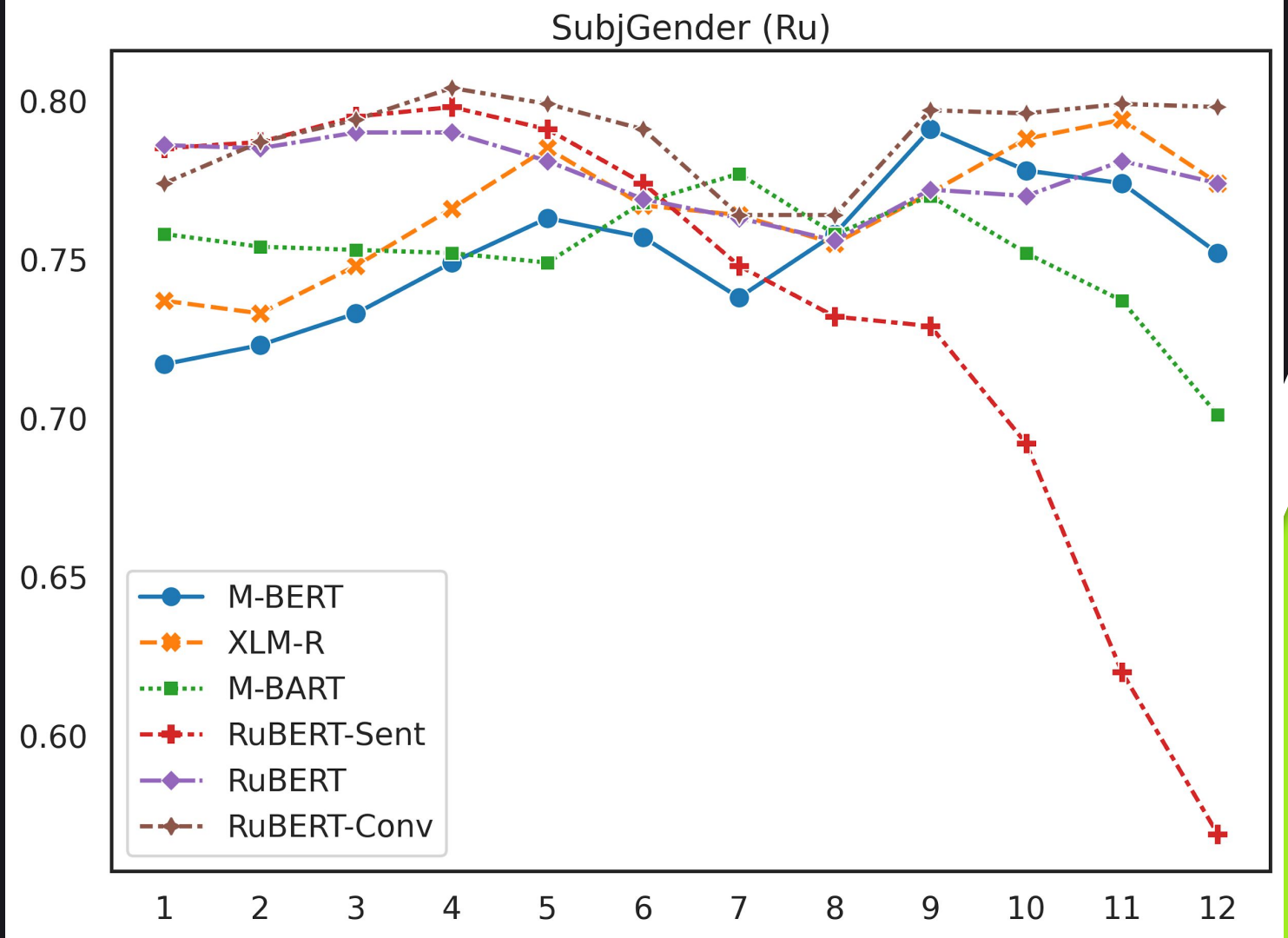
How do we distinguish the good embeddings from the bad?
Probing!

Let's use annotated Russian sentences and get their embeddings from different layers from the model

+ a simple classifier on top
+ sentence annotation on embeddings
+ bad classification quality = no info in embeddings = bad embeddings


TreeDepth (Ru)
0.475
0.450
0.425
0.400
0.375
0.350
0.325
0.300
1
2
3
4
5
6
7
8
9
10
11
12
M-BERT
XLM-R
M-BART
RuBERT-Sent
RuBERT
RuBERT-Conv


SubjNumber (Ru)
0.94
0.92
0.90
0.88
0.86
0.84
0.82
1
2
3
4
5
6
7
8
9
10
11
12
M-BERT
XLM-R
M-BART
RuBERT-Sent
RuBERT
RuBERT-Conv


SubjGender (Ru)
0.80
0.75
0.70
0.65
0.60
1
2
3
4
5
6
7
8
9
10
11
12
M-BERT
XLM-R
M-BART
RuBERT-Sent
RuBERT
RuBERT-Conv

github.com/RussianNLP/rusenteval

RussianNLP / **rusenteval**

Unwatch

Code | Issues 1 | Pull requests | Actions | Projects | Wiki | Security | Insights | Settings

main | 1 branch | 0 tags | Go to file | Add file | Code

vmkhlv Update README.md 7868207 5 hours ago 15 commits

data	added tasks	4 months ago
images	Add files via upload	19 hours ago
probing	added rusenteval code	6 days ago
README.md	Update README.md	5 hours ago
install_tools.sh	added rusenteval code	6 days ago
requirements.txt	added rusenteval code	6 days ago

README.md

RuSentEval

Linguistic Source, Encoder Force!

RuSentEval is an evaluation toolkit for sentence embeddings for Russian.

In this repo you can find the data and scripts to run an evaluation of the quality of sentence embeddings.

About

No description, website, or topics provided.

Readme

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Contributors 3

TatianaShavrina Tatiana Shavrina

vmkhlv Vlad Mikhailov

artemovae Katya Artemova

MOROCCO framework

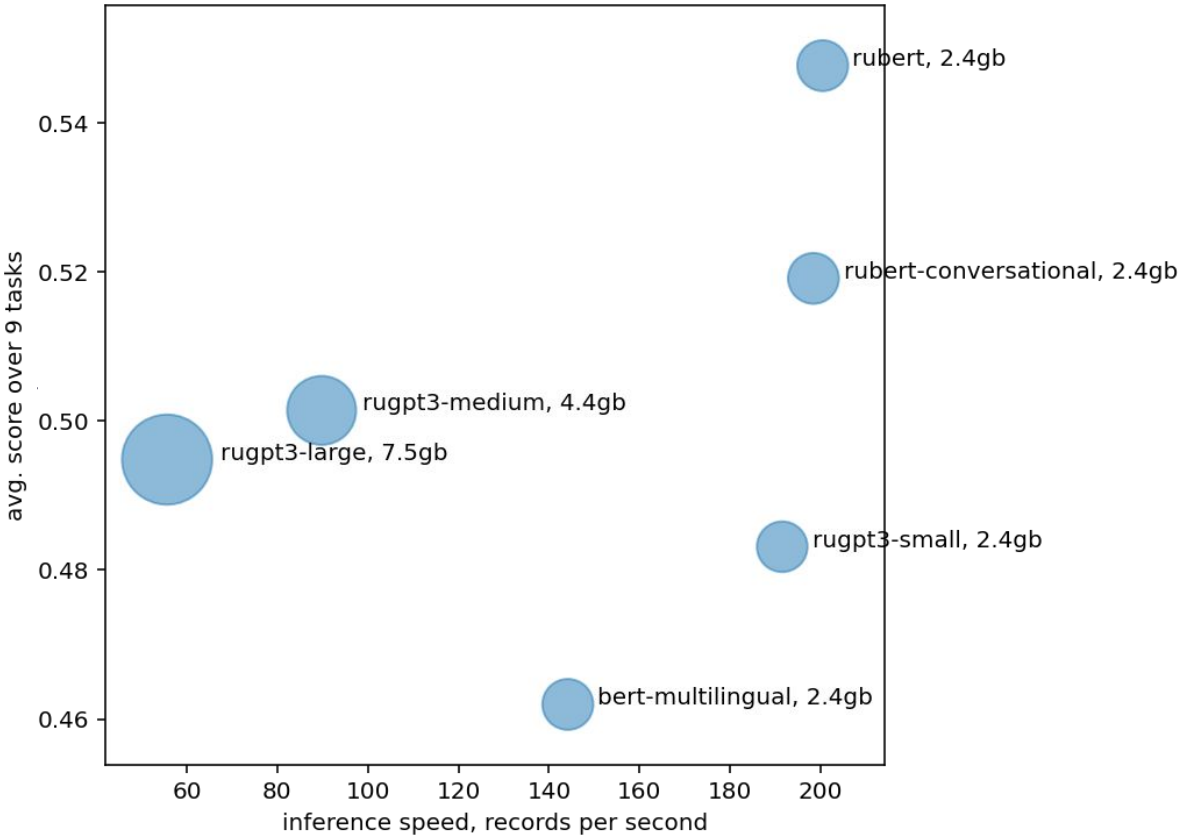
MOdel ResOurCe COnsumption

MOROCCO idea

Lets evaluate models by GPU RAM usage + inference speed + Russian SuperGLUE score

- Model results on GLUE are unstable, depend on random seed
- Smaller models have higher inference speed. `rugpt3-small` processes ~200 records per second while `rugpt3-large` — ~60 records/second.
- `bert-multilingual` is a bit slower then `rubert*` due to worse Russian tokenizer. `bert-multilingual` splits text into more tokens, has to process larger batches.
- It is common that larger models show higher score but in our case `rugpt3-medium`, `rugpt3-large` perform worse then smaller `rubert*` models.
- `rugpt3-large` has more parameters then `rugpt3-medium` but is currently trained for less time and has lower score.

Russian Models by inference speed and performance



Спасибо за attention!

AGI tasks: github.com/RussianNLP/RussianSuperGLUE
RAM & speed: github.com/RussianNLP/MOROCCO
Probing: github.com/RussianNLP/rusenteval

@rybolos SberDevices, HSE University, Huawei