

# Intro to DeepPavlov. Multi-Task Learning.

Dmitry Karpov

Moscow Institute of Physics and Technology



Библиотека DeepPavlov

Как использовать библиотеку

Многозадачное обучение

Многозадачное обучение - исследования переноса знаний на диалоговых наборах данных

## Определение

Открытая библиотека обработки естественного языка DeepPavlov<sup>1</sup> предназначена для использования многоязычных и языко-специфичных NLP моделей как самостоятельно, так и в рамках диалоговых систем.

---

<sup>1</sup>Mikhail Burtsev и др. «DeepPavlov: An Open Source Library for Conversational AI». В: *NIPS*. 2018. URL:

<https://openreview.net/pdf?id=BJzyCF6Vn7>.

# Особенности библиотеки

- ▶ DeepPavlov основан на json конфигурационных файлах моделей
- ▶ DeepPavlov использует transformers от HuggingFace
- ▶ DeepPavlov поддерживает REST API и может использоваться в качестве сервиса
- ▶ DeepPavlov распространяется под лицензией Apache 2.0

# Ссылки на DeepPavlov

- ▶ DeepPavlov [Demo](#)[Click]
- ▶ DeepPavlov [GitHub](#)[6k Stars]
- ▶ DeepPavlov [Forum](#)
- ▶ DeepPavlov [Documentation](#)

# Поддерживаемые модели

- ▶ Named-Entity Recognition
- ▶ Sentiment and Emotion Classification
- ▶ Question Answering
- ▶ GLUE/SuperGLUE
- ▶ Russian SuperGLUE
- ▶ Few-shot models

# Как использовать библиотеку



Пример

# Конфигурационные файлы для поддерживаемых моделей

Russian Paraphraser Russian NER English NER Russian SQUAD  
English SQUAD Russian Sentiment English Sentiment Russian  
relation ranking Russian ODQA English ODQA Russian KBQA  
English KBQA



# Многозадачное обучение - зачем оно нужно

- ▶ Экономия вычислительных ресурсов
- ▶ Рост качества на схожих задачах

# Многозадачное обучение - как оно реализовано (энкодер-агностичные модели)

В 2 словах - отдельный задаче-специфичный линейный слой на каждую задачу, который применяется к выходу энкодера.

Достоинства архитектуры:

- ▶ Вычислительная и архитектурная простота
- ▶ Расширяемость на различные типы задач
- ▶ Не требует псевдоразметки
- ▶ Можно быстро заменить энкодер

Архитектура интегрирована в open-source библиотеку DeepPavlov. [Пример](#)

# Многозадачное обучение - результаты на GLUE

htbp

**Таблица:** Метрики многозадачной энкодер-агностичной модели для набора задач GLUE. M.Corr означает корреляцию Мэттью, P/S означает корреляцию Пирсона-Спирмена, Acc точность, F1 - макро-F1. Режим S означает однозадачные модели, режим M означает многозадачные модели. Размер означает размер тренировочного набора данных.

Модель	Режим	Среднее	CoLA	SST-2	MRPC	STS-B	QQP	MNLI	QNLI	RTE	AX
		Размер	8.6k	67.3k	2.5k	5.7k	363.8k	392.7k	104.7k	2.5k	как у MNLI
		метрика	M.Corr	Acc	F1/Acc	P/S Corr	F1/Acc	Acc (m/mm)	Acc	Acc	M.Corr
Человек	-	87.1	66.4	97.8	86.3/80.8	92.7/92.6	59.5/80.4	92.0/92.8	91.2	93.6	-
<i>distilbert</i>	S	73.3	<b>42.4</b>	<b>92.1</b>	<b>85.6/80.3</b>	<b>78.8/76.8</b>	<b>69.5/88.5</b>	<b>81.3/80.8</b>	<b>87.5</b>	52.1	29.9
	M	<b>74.5</b>	36.0	91.0	<b>85.7/79.9</b>	<b>82.6/81.6</b>	68.4/87.4	80.4/80.3	86.0	<b>69.5</b>	<b>30.1</b>
<i>bert</i>	S	77.3	<b>53.7</b>	<b>93.2</b>	<b>87.7/82.8</b>	<b>83.8/82.2</b>	<b>70.3/88.9</b>	<b>83.8/83.1</b>	<b>90.6</b>	62.1	32.1
	M	<b>77.8</b>	45.8	92.9	86.8/82.2	<b>85.3/84.7</b>	70.2/88.6	83.5/82.6	90.1	<b>74.5</b>	<b>32.8</b>
<i>bert-large</i>	S	<b>79.5</b>	<b>59.2</b>	<b>94.9</b>	85.0/80.6	<b>85.8/84.5</b>	70.5/89.1	<b>86.7/85.6</b>	92.2	70.1	<b>39.4</b>
	M	<b>79.5</b>	50.8	94.1	<b>87.3/82.8</b>	83.8/83.9	<b>71.0/89.2</b>	85.9/85.0	<b>92.4</b>	<b>78.5</b>	38.5

# Энкодер-агностичные модели: данные, сэмплирование

Принцип подбора данных:

- ▶ Разговорные задачи
- ▶ Совпадающие классы для английского и русского языка

Сэмплирование примеров на каждом этапе обучения - батч из каждого набора данных с вероятностью пропорционально размеру (plain sampling).

## Энкодер-агностичные модели: данные

- ▶ Для классификации **эмоций** – русскоязычный набор данных CEDR, собранный из различных интернет-источников, и англоязычный набор данных go\_emotions, собранный из комментариев на ресурсе «Реддит». Использовалось семь типов эмоций по Экману – ярость, страх, грусть, удовольствие, удивление, отвращение, нейтральная.
- ▶ Для классификации **тональности** – англоязычный набор данных DynaSent(r1), состоящий из предложений, возникающих в диалогах, и русскоязычный набор данных RuReviews, состоящий из отзывов крупного российского электронного магазина. Использовалось три класса – положительный, отрицательный, нейтральный.

## Энкодер-агностичные модели: данные

- ▶ Для классификации **токсичности** – русскоязычный набор комментариев с ресурса «Двач» (RuToxic) и англоязычный набор комментариев из Википедии (Wiki Talk).  
Использовалось два класса – токсичный и не токсичный.
- ▶ Для классификации **тем** и классификации **интентов** – набор данных MASSIVE, состоящий из обращенных к диалоговой системе фраз пользователей. Набор существует и использовался как в англоязычном, так и в русскоязычном варианте. Каждая фраза из набора принадлежит к одной из 60 тем и к одному из 18 интентов.

# Энкодер-агностичные модели: сравнение с однозадачными, английский язык

**Таблица:** Метрики англоязычных моделей (точность/макро-F1) для пяти англоязычных диалоговых задач. Режим S означает однозадачные модели, режим M означает многозадачные модели. Усреднено по трем запускам.

Модель	Режим	Среднее	Эмоции 39.4k	Тональность 80.5k	Токсичность 127.6k	Интененты 11.5k	Темы 11.5k
<i>distilbert</i>	S	<b>82.9</b>	<b>70.3</b>	74.7	91.5	<b>87.4</b>	<b>91.0</b>
	M	82.1	67.7	<b>75.2</b>	90.6	86.3	90.8
<i>bert</i>	S	<b>83.9</b>	<b>71.2</b>	76.1	<b>93.2</b>	<b>87.9</b>	<b>91.3</b>
	M	83.0	69.0	<b>76.5</b>	91.4	87.1	91.2
<i>bert-large</i>	S	<b>84.7</b>	<b>70.9</b>	<b>80.5</b>	<b>92.1</b>	<b>88.4</b>	91.3
	M	83.6	69.0	79.0	91.3	87.3	<b>91.3</b>

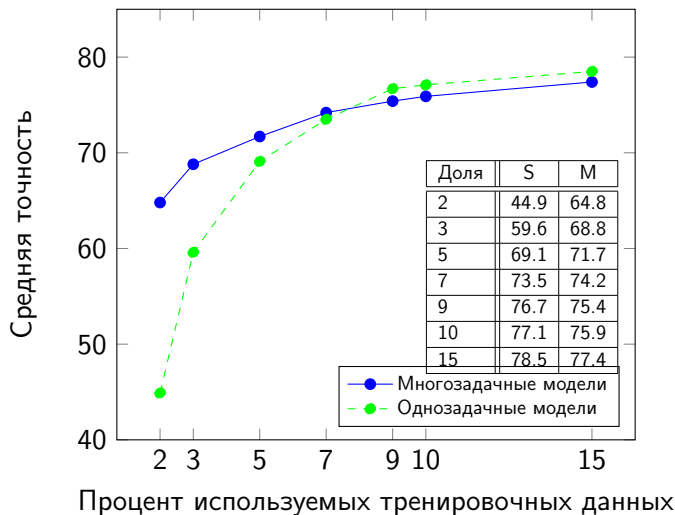
# Энкодер-агностичные модели: сравнение с однозадачными, русский язык

**Таблица:** Метрики русскоязычных моделей (точность/f1 macro) для пяти диалоговых задач. Режим S означает однозадачные модели, режим M означает многозадачные модели. Усреднено по трем запускам.

Модель	Режим	Среднее	Эмоции 6.5k	Тональность 82.6k	Токсичность 93.3k	Интененты 11.5k	Темы 11.5k
<i>distilrubert</i>	S	86.9	82.2	77.9	97.1	86.7	90.4
	M	86.3	81.0	77.7	96.9	85.2	90.7
<i>rubert</i>	S	86.5	80.9	78.0	97.2	86.2	90.0
	M	86.2	80.5	77.6	96.8	85.3	90.5



## Энкодер-агностичные модели - эффект уменьшения размера выборки, данные на английском языке

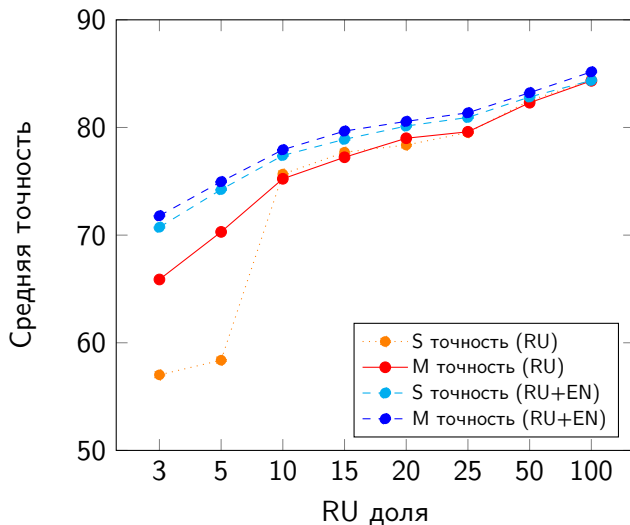


# Энкодер-агностичные модели - эффект уменьшения размера выборки, данные на русском языке

**Таблица:** Средняя точность многозадачных моделей в зависимости от того, на какой доле русскоязычных данных они обучались, и того, добавлялись ли к ним англоязычные данные, для многоязычного distilbert.

RU доля	S RU	M RU	S RU+EN	M RU+EN
3	57.0	<b>65.9</b>	<b>71.8</b>	70.7
5	58.4	<b>70.3</b>	<b>75.0</b>	74.2
10	<b>75.7</b>	75.2	<b>77.9</b>	77.4
15	<b>77.7</b>	77.2	<b>79.7</b>	78.9
20	78.4	<b>79.0</b>	<b>80.6</b>	80.1
25	79.5	<b>79.6</b>	<b>81.4</b>	80.9
50	<b>82.5</b>	82.3	<b>83.2</b>	82.8
100	<b>84.4</b>	84.3	<b>85.2</b>	84.4

## Энкодер-агностичные модели - эффект уменьшения размера выборки, данные на русском языке



# Энкодер-агностичные модели: выводы

1. Многозадачные энкодер-агностичные модели - почти как однозадачные. Если для какой-то задачи данных мало, но для другой похожей задачи их много - то даже лучше.
2. Если данных становится очень мало, то многозадачные модели становятся сильно лучше однозадачных. Опять же, зависит от размера данных для задачи.
3. Добавление английских данных к русским - улучшает метрики, чем меньше русских данных - тем сильнее (до нескольких процентов). Это верно и для однозадачных моделей, при любом языке валидации.

Спасибо за внимание