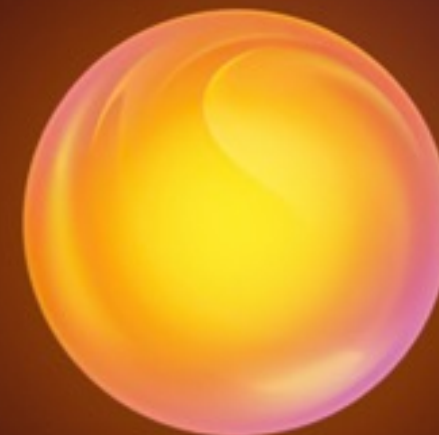


# Dusha: самый большой открытый датасет для распознавания эмоций в устной речи на русском языке

Кондратенко Владимир

ML Engineer, SberDevices



## О спикере

- Делаю распознавание эмоций в **SaluteSpeech**
- Разрабатываю модели и алгоритмы в **SaluteSpeech Insights**
- Один из авторов **Dusha**

# О чем поговорим

1. Почему интересно
2. Что существует
3. Что побудило
4. Как собирали
5. Что получилось
6. Что в репозитории
7. Q&A

Поехали

# Почему наш датасет интересен

1

Интересная задача

2

Самый большой  
открытый датасет\*.  
На русском языке

3

Новый бенчмарк

\*Для задачи Speech Emotion Recognition (SER)

# Существующие датасеты

1

**IEMOCAP**

5к семплов, 12 часов, 10 актеров

2

**MSP Podcast**

62к семплов, 100 часов, ~150 спикеров

3

**RUSLANA**

6к семплов, 61 спикер

4

**RAMAS**

581 семпл, 7 часов, 10 актеров

5

**MURCO**

много семплов, эмоции не размечались

# Проблемы существующих датасетов

- Объем данных
- Количество спикеров
- Реалистичность эмоций и их распределения
- Качество разметки

Для чего мы выложили датасет в открытый доступ

1

Предлагаем бенчмарк для  
Speech Emotion Recognition

2

Помогаем развитию  
научной области



# Сбор датасета

# Из чего состоит Dusha

1

## Crowd

201к семплов, 255 часов, 2к спикеров

Наигранные эмоции, «концентрированные»

2

## Podcast

102к семплов, 90 часов, 6к спикеров

Реальные эмоции, реальный баланс

## Crowd. Подготовка текстов для озвучки

- Тексты похожи на реальные
- Текст и эмоция для озвучки соотносятся друг с другом

## Crowd. Подготовка текстов для озвучки

1. Соберем тексты похожие на реальные
2. Оценим эмоцию для каждого текста

# Crowd. Оценка эмоций в текстах для озвучки

## Позитив

Хвастается успехами («я получил пятерку»), вокруг происходит что-то хорошее («у нас классная погода», хвалит кого-то («он очень умный»), смех (есть «ха-ха» в тексте), поздравление («с Новым годом»), что-то понравилось («эта песня мне нравится»), хочу дружить/давай дружить.

## Грусть

Скучно/грустно/печально или плохое настроение, текст, требующий сочувствия («заболел хомячок», «что-то с ногой, хромаю»), что-то не получается («не могу уровень пройти»), жалуется на работу, погоду и т. д., («замучали меня», «я не могу на улицу выйти уже две недели»).

## Совсем непонятно

Так предлагалось помечать записи, если текст совсем несвязный

## Злость / Раздражение

Нецензурная лексика/оскорбления/ угрозы («я тебя сломаю», «слышь, я не буду это делать»)/раздражение («ты врешь мне», «я сказала выйти из меню»)/что-то не работает, не включается.

## Нейтраль

Если текст не подходит ни под одну из всех остальных эмоций/люди запрашивают музыку, фильм или какой-то другой контент.

## Crowd. Оценка эмоций в текстах для озвучки

1. Разметили тестовый сет
2. Подобрали пороги у BERT
3. Отобрали тексты в которых BERT достаточно уверен
4. Получили пары (текст, эмоция)

# Crowd. Озвучка текстов

Группа 1 озвучила тексты с интонацией или нейтралью:

## Позитив

Текст с такой эмоцией необходимо произносить с улыбкой или смехом, стараться делать выраженные ударения на словах, подчеркивающих позитив

## Злость / Раздражение

Надо злиться, кричать, негодовать

## Нейтраль

Прочитать фразу максимально «плоско», не выражая никаких эмоций

## Грусть

Произносить текст с грустью, как будто у тебя был тяжелый рабочий день или приболел хомячок

## Crowd. Озвучка текстов

1. Группа 1 озвучила тексты
2. Группа 2 оценила эмоцию полученного аудио
3. Отобрали только совпавшие



## Crowd. Резюме

1. Отобрали тексты
2. Получили пары (текст, эмоция)
3. Озвучили тексты с соответствующей эмоцией или нейтралью
4. Распознали эмоции в «наговоренных» аудио
5. Оставили только совпадающие

# Podcast

- Нарезки подкастов на русском языке
- От 1 до 5 секунд
- Игры, IT, отношения, политика
- 6240 разных спикеров

# Podcast. Разметка аудио

## Позитив

Текст произнесён с улыбкой, смехом, восхищением, игривым тоном, или есть выраженные ударения на словах, подчеркивающих позитив (например, «вот ты красАвец!», «АХ, как здорово»)

## Нейтраль

Голос ровный, никакой характерной эмоции в голосе нет. При этом даже, если текст явно негативный (например, «да как ты надоел») или позитивный (например, «какой ты молодец»), в голосе эта эмоция не выражается, тогда необходимо ставить нейтраль.

## Совсем непонятно

Если запись слишком тихая, шипение, стук и никаких голосов

## Злость / Раздражение

Текст произнесён со злостью, раздражением, или пользователь кричит, говорит сквозь зубы, или есть выраженные ударения на словах, подчеркивающих негатив (например, «да пошёл ты!», «как ты задолБАЛ»)

## Грусть

Текст произнесён с грустью/тоской, потухшим голосом

Что получилось

## Статистики

Domain	Number of Files	Duration (Hr.)	Unique Speakers	Train Labels	Test Labels	Labels per Record
Crowd	201 850	255.7	2068	906 953	79 088	4.89
Podcast	102 113	90.9	6240	645 813	83 812	7.14
Total	303 963	346.6	8308	1 552 766	162 900	5.85

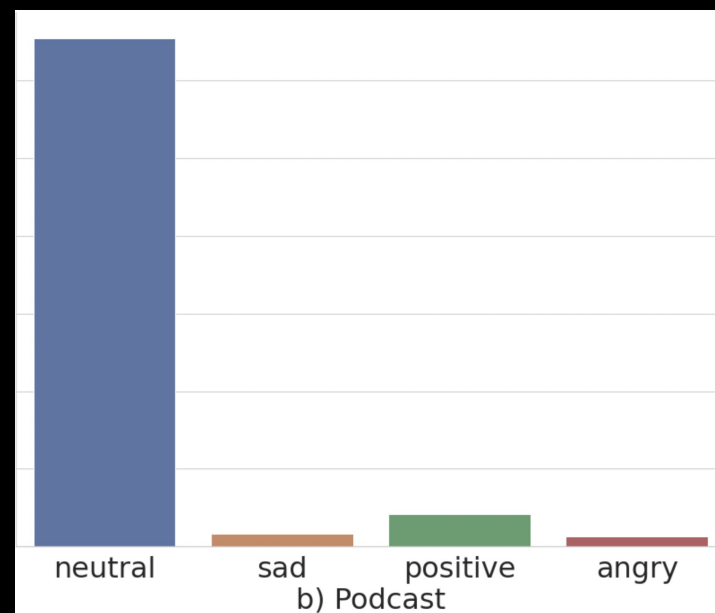
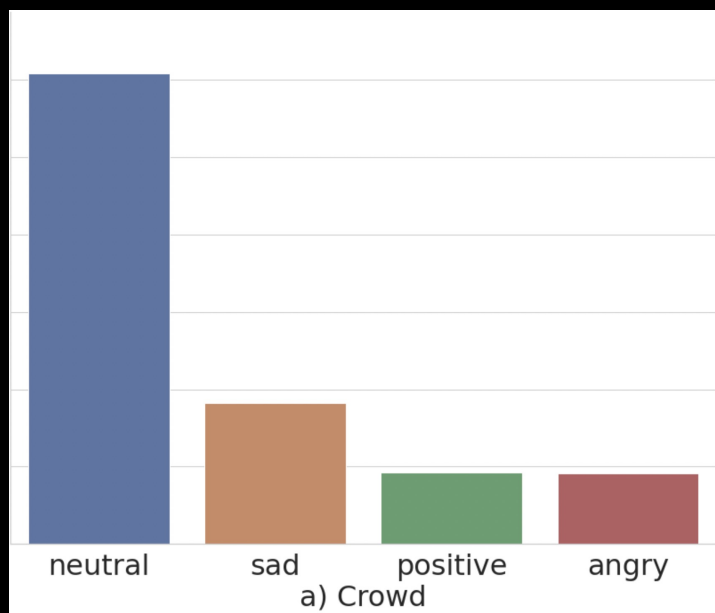
Domain	Training files and hours		Test files and hours	
Crowd	150 352	188 h. 44 min.	14 035	18 h. 29 min.
Podcast	798 25	71 h. 23 min.	10 637	09 h. 25 min.
Total	230 177	260 h. 07 min.	24 672	27 h. 54 min.

*Amount of Data After Aggregation Mechanism using Dawid-Skene algorithm with threshold 0.9*

# Распределение эмоций

Domain	Pos	Sad	Ang	Neu	0th	Total
Crowd	15446	23316	17120	106850	1655	164387
Podcast	5909	1170	2057	81104	222	90462

*Emotion Files Distribution After Aggregation Mechanism using Dawid-Skene algorithm with threshold 0.9*



# Что в репозитории

1

## Бенчмарк

Сарегистрированная разметка

Train / test split

Предсчитанные фичи

2

## Разметка до агрегации

Полная информация о разметке до  
фильтрации

Код для воспроизводимости агрегации

3

## Код бейзлайна

Код обучения бейзлайна

Код для подсчета метрик

Docker image

# Резюме

## — Собрали большой датасет

1. Crowd с «наговоренными» эмоциями
2. Podcast с реальными эмоциями

## — Дополнительно опубликовали «сырую» разметку, код бейзлайна

## — Можно использовать по-разному

- Бенчмарк Speech Emotion Recognition
- Исследования Speech Emotion Recognition
- Исследования различных алгоритмов агрегации



## Авторы датасета

- Артем Соколов
- Федор Минькин
- Никита Савушкин
- Николай Карпов
- Олег Кутузов
- Владимир Кондратенко

# Контакты

 @username27

[vakondratenko@sberbank.ru](mailto:vakondratenko@sberbank.ru)

<https://github.com/salute-developers/golos/tree/master/dusha>



