Практическая работа №2

Потоковая аналитика эмоциональных оттенков речи человека

Цель работы:

Выполнение практической работы направлено на изучение современных программных средств определения эмоционального фона речи человека в голосе и в тексте.

Порядок работы:

1. Определение эмоционального фона речи человека в текстовой форме с использованием модуля *Text* библиотеки *Aniemore* для *Python*. Проверка модели *rubert-tiny-emotion-russian-cedr-m7*. Программа определила эмоциональный фон как «нейтральный» с точностью 38%, как показано на рисунке 1.

```
# @title Imports and variables settings
from rich import print  # for better view
import torch
from aniemore.recognizers.text import TextRecognizer
from aniemore.models import HuggingFaceModel

model = HuggingFaceModel.Text.Bert_Tiny
device = 'cuda' if torch.cuda.is_available() else 'cpu'

# tr - acronym for TextRecognizer from two first capital letters
tr = TextRecognizer(model=model, device=device)
```

Recognize: single text - single label

```
[20] # @title Recognize: single text - single label
text = 'если ты не прекратишь, я тебя убью!'
result = tr.recognize(text, return_single_label=True), tr.recognize(text, return_single_label=False)
print(result)
```

```
(
    'neutral',
    {
        'anger': 0.33996066451072693,
        'disgust': 0.025116577744483948,
        'enthusiasm': 0.018978049978613853,
        'fear': 0.1447315365076065,
        'happiness': 0.012643322348594666,
        'neutral': 0.3827205002307892,
        'sadness': 0.2320241630077362
    }
)
```

Рисунок 1 - Проверка модели архитектуры *Bert_Tiny*

2. Проверка модели *rubert-tiny2-russian-emotion-detection*. Программа определила эмоциональный фон как «гневный» с точностью 80%, как показано на рисунке 2.

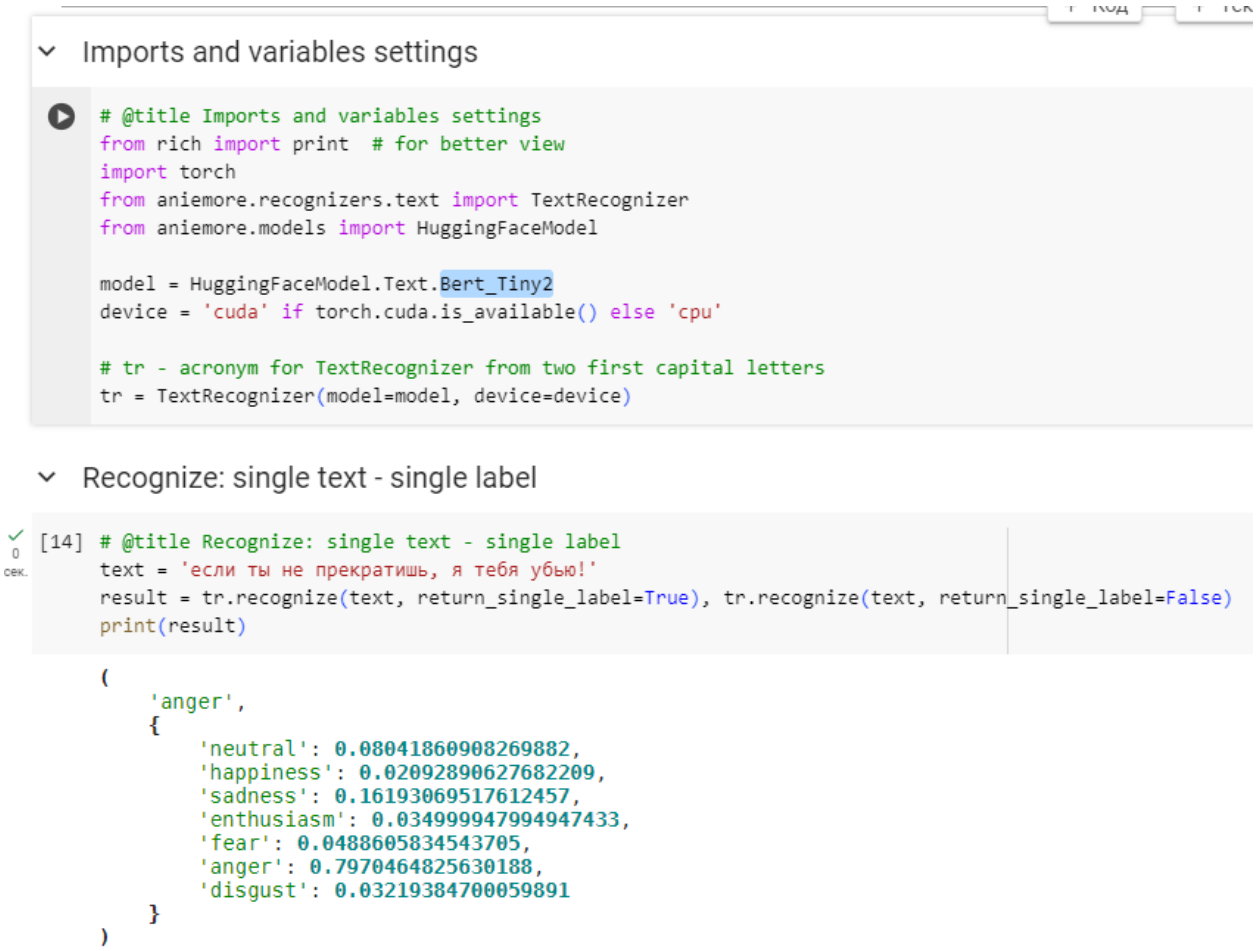

Рисунок 2 - Проверка модели архитектуры *Bert_Tiny2*

3. Проверка модели *rubert-large-emotion-russian-cedr-m7*. Программа определила эмоциональный фон как «гневный» с точностью 78%, как показано на рисунке 3.

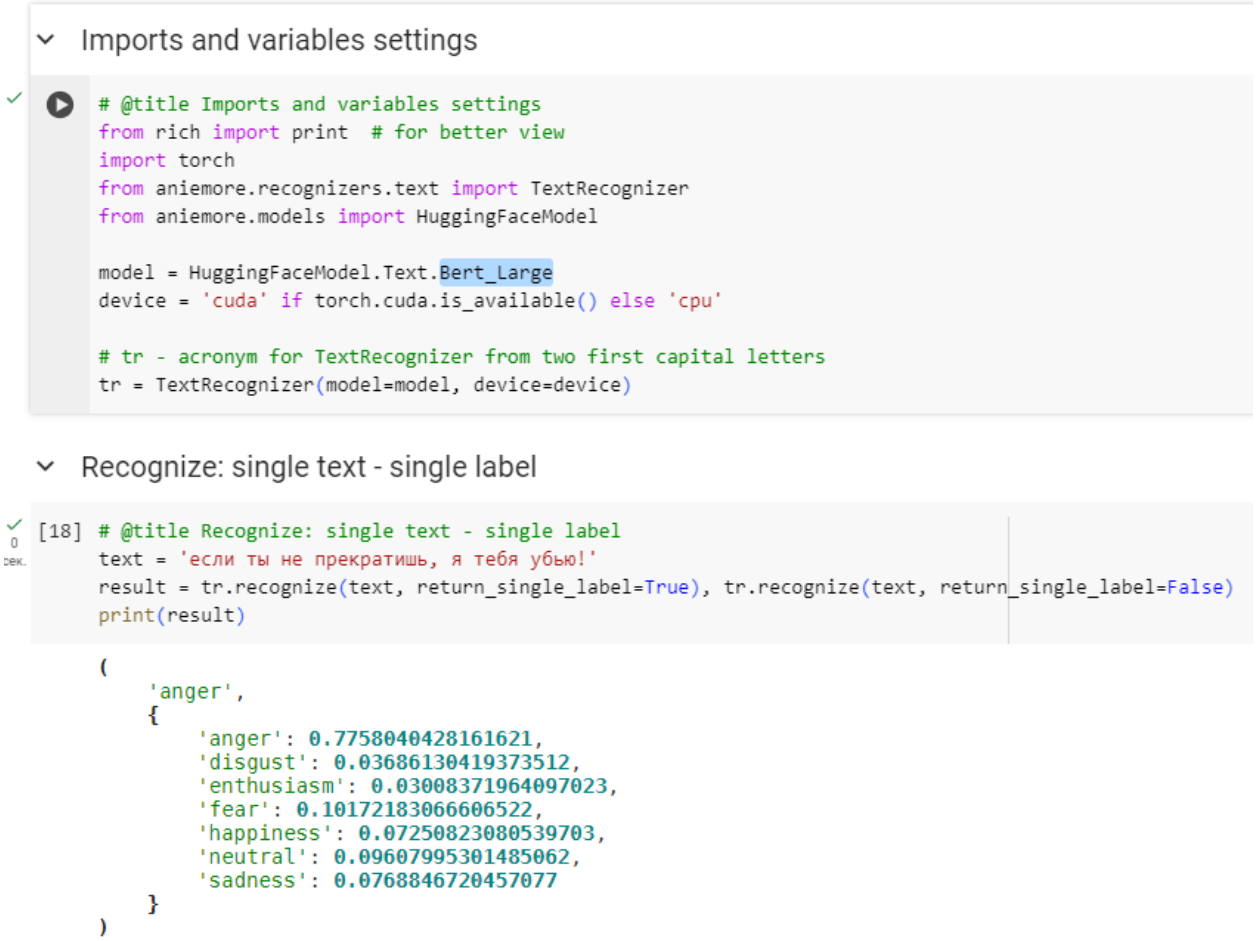

Рисунок 3 - Проверка модели архитектуры *Bert_Large*

4. Проверка модели *rubert-base-emotion-russian-cedr-m7*. Программа определила эмоциональный фон как «гневный» с точностью 43%, как показано на рисунке 4.

Imports and variables settings

```
# @title Imports and variables settings
from rich import print  # for better view
import torch
from aniemore.recognizers.text import TextRecognizer
from aniemore.models import HuggingFaceModel

model = HuggingFaceModel.Text.Bert_Base
device = 'cuda' if torch.cuda.is_available() else 'cpu'

# tr - acronym for TextRecognizer from two first capital letters
tr = TextRecognizer(model=model, device=device)
```

Recognize: single text - single label

```
[16] # @title Recognize: single text - single label
text = 'если ты не прекратишь, я тебя убью!'
result = tr.recognize(text, return_single_label=True), tr.recognize(text, return_single_label=False)
print(result)
```

```
(
    'anger',
    {
        'anger': 0.4286688268184662,
        'disgust': 0.01858402229845524,
        'enthusiasm': 0.028649266809225082,
        'fear': 0.06029261648654938,
        'happiness': 0.04082256928086281,
        'neutral': 0.3020907938480377,
        'sadness': 0.09338401257991791
    }
)
```

Рисунок 4 - Проверка модели архитектуры *Bert_Base*

5. Подготовка к определению эмоционального фона речи человека в голосовой форме с использованием модуля *Voice* библиотеки *Aniemore* для *Python*. Загрузка файла голосовой записи представлена на рисунке 5.

Load our test voicel file

```
[21] # @title Load our test voicel file
import wget
from pathlib import Path
voice_url = "https://github.com/aniemore/Aniemore/blob/master/tests/aniemore/recognizers/src/
voices_path: Path = Path('my_voice.ogg')

wget.download(voice_url, str('my_voice.ogg'))
```

```
'my_voice.ogg'
```

Рисунок 5 - Загрузка звукового файла

6. Проверка модели *wavlm-emotion-russian-resd*. Программа определила эмоциональный фон как «отвращенный» с точностью 71%, как показано на рисунке 6.

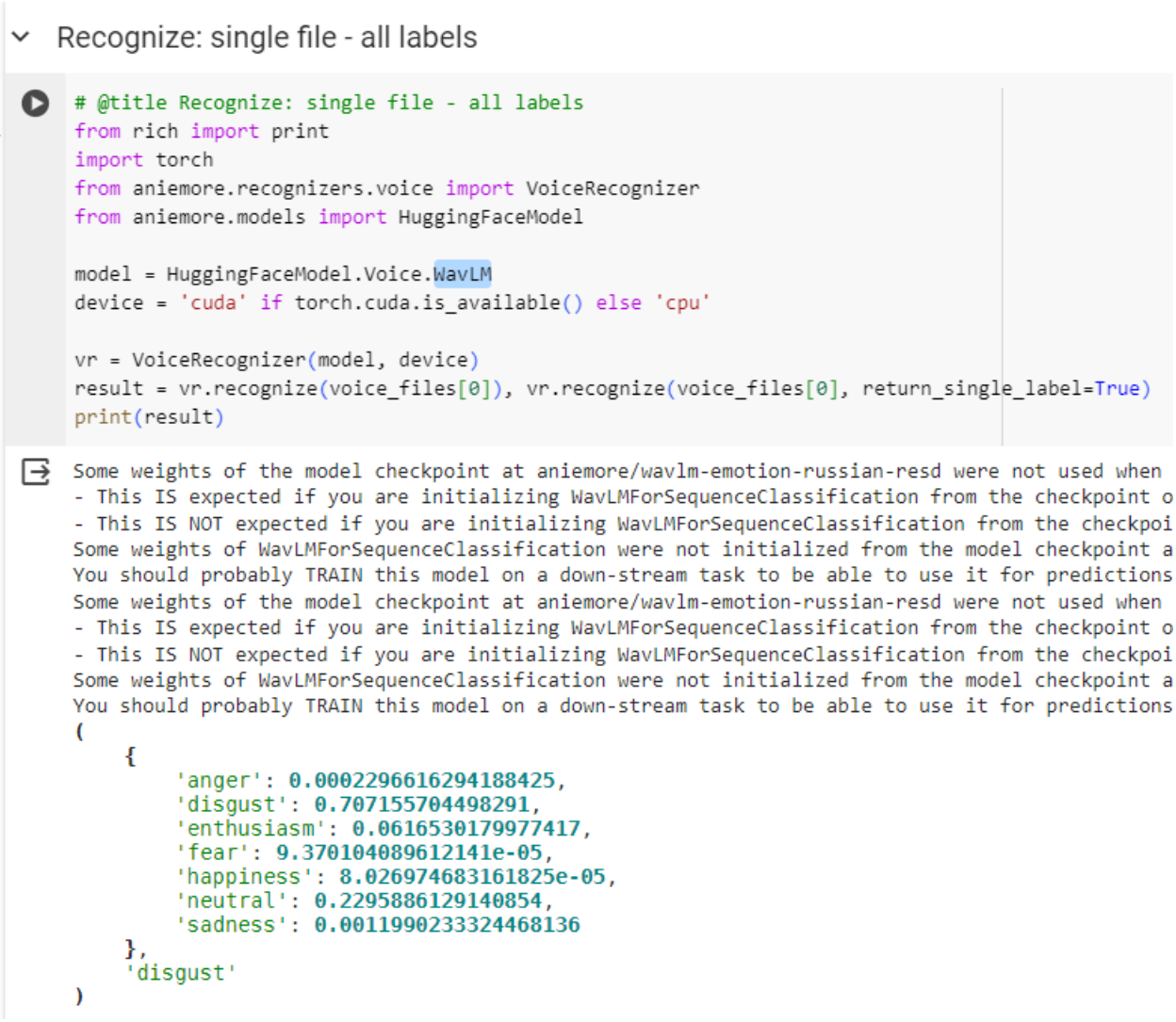

Рисунок 6 - Проверка модели архитектуры *WavLM*

7. Проверка модели *wav2vec2-xlsr-53-russian-emotion-recognition*. Программа определила эмоциональный фон как «нейтральный» с точностью 99%, как показано на рисунке 7.

```
# @title Recognize: single file - all labels
from rich import print
import torch
from aniemore.recognizers.voice import VoiceRecognizer
from aniemore.models import HuggingFaceModel

model = HuggingFaceModel.Voice.Wav2Vec2
device = 'cuda' if torch.cuda.is_available() else 'cpu'

vr = VoiceRecognizer(model, device)
result = vr.recognize(voice_files[0]), vr.recognize(voice_files[0], return_single_label=True)
print(result)
```

```
Some weights of the model checkpoint at aniemore/wav2vec2-emotion-russian-resd were not used w
- This IS expected if you are initializing Wav2Vec2ForSequenceClassification from the checkpoi
- This IS NOT expected if you are initializing Wav2Vec2ForSequenceClassification from the chec
Some weights of Wav2Vec2ForSequenceClassification were not initialized from the model checkpoi
You should probably TRAIN this model on a down-stream task to be able to use it for prediction
Some weights of the model checkpoint at aniemore/wav2vec2-emotion-russian-resd were not used w
- This IS expected if you are initializing Wav2Vec2ForSequenceClassification from the checkpoi
- This IS NOT expected if you are initializing Wav2Vec2ForSequenceClassification from the chec
Some weights of Wav2Vec2ForSequenceClassification were not initialized from the model checkpoi
You should probably TRAIN this model on a down-stream task to be able to use it for prediction
(
    {
        'anger': 1.5641922800568864e-05,
        'disgust': 0.0001745589543133974,
        'enthusiasm': 0.0008272648556157947,
        'fear': 4.06472872782615e-06,
        'happiness': 1.4103216017247178e-05,
        'neutral': 0.9977772831916809,
        'sadness': 0.0011870539747178555
    },
    'neutral'
)
```

Рисунок 7 - Проверка модели архитектуры *Wav2Vec2*

8. Проверка модели *hubert-emotion-russian-resd*. Программа определила эмоциональный фон как «воодушевленный» с точностью 92%, как показано на рисунке 8.

```python
# @title Recognize: single file - all labels
from rich import print
import torch
from aniemore.recognizers.voice import VoiceRecognizer
from aniemore.models import HuggingFaceModel

model = HuggingFaceModel.Voice.Hubert
device = 'cuda' if torch.cuda.is_available() else 'cpu'

vr = VoiceRecognizer(model, device)
result = vr.recognize(voice_files[0]), vr.recognize(voice_files[0], return_single_label=True)
print(result)
```

```
Some weights of the model checkpoint at aniemore/hubert-emotion-russian-resd were not used when
- This IS expected if you are initializing HubertForSequenceClassification from the checkpoint o
- This IS NOT expected if you are initializing HubertForSequenceClassification from the checkpoi
Some weights of HubertForSequenceClassification were not initialized from the model checkpoint a
You should probably TRAIN this model on a down-stream task to be able to use it for predictions
Some weights of the model checkpoint at aniemore/hubert-emotion-russian-resd were not used when
- This IS expected if you are initializing HubertForSequenceClassification from the checkpoint o
- This IS NOT expected if you are initializing HubertForSequenceClassification from the checkpoi
Some weights of HubertForSequenceClassification were not initialized from the model checkpoint a
You should probably TRAIN this model on a down-stream task to be able to use it for predictions
(
    {
        'anger': 1.8258066120324656e-05,
        'disgust': 0.03384198993444443,
        'enthusiasm': 0.9230761528015137,
        'fear': 6.742441473761573e-05,
        'happiness': 7.468734111171216e-05,
        'neutral': 0.041868288069963455,
        'sadness': 0.0010530701838433743
    },
    'enthusiasm'
)
```

Рисунок 8 - Проверка модели архитектуры *Hubert*

9. Проверка модели *unispeech-sat-emotion-russian-resd*. Программа определила эмоциональный фон как «воодушевленный» с точностью 64%, как показано на рисунке 9.

```
# @title Recognize: single file - all labels
from rich import print
import torch
from aniemore.recognizers.voice import VoiceRecognizer
from aniemore.models import HuggingFaceModel

model = HuggingFaceModel.Voice.UniSpeech
device = 'cuda' if torch.cuda.is_available() else 'cpu'

vr = VoiceRecognizer(model, device)
result = vr.recognize(voice_files[0]), vr.recognize(voice_files[0], return_single_label=True)
print(result)
```

```
Some weights of the model checkpoint at aniemore/unispeech-sat-emotion-russian-resd were not use
- This IS expected if you are initializing UniSpeechSatForSequenceClassification from the checkp
- This IS NOT expected if you are initializing UniSpeechSatForSequenceClassification from the ch
Some weights of UniSpeechSatForSequenceClassification were not initialized from the model checkp
You should probably TRAIN this model on a down-stream task to be able to use it for predictions
Some weights of the model checkpoint at aniemore/unispeech-sat-emotion-russian-resd were not use
- This IS expected if you are initializing UniSpeechSatForSequenceClassification from the checkp
- This IS NOT expected if you are initializing UniSpeechSatForSequenceClassification from the ch
Some weights of UniSpeechSatForSequenceClassification were not initialized from the model checkp
You should probably TRAIN this model on a down-stream task to be able to use it for predictions
(
    {
        'anger': 0.00022612408793065697,
        'disgust': 0.20026256144046783,
        'enthusiasm': 0.637061595916748,
        'fear': 0.00029982629348523915,
        'happiness': 0.00013136488269083202,
        'neutral': 0.14902694523334503,
        'sadness': 0.012991590425372124
    },
    'enthusiasm'
)
```

Рисунок 9 - Проверка модели архитектуры *UniSpeech*

10. Результаты исследований приведены в таблице 1.

Таблица 1 - Сравнение точности

Модель	Заявленная точность	Подтвержденная точность
Голосовые модели		
wav2vec2-xlsr-53-russian-emotion-recognition	73%	99%
wav2vec2-emotion-russian-resd	75%	99%
wavlm-emotion-russian-resd	82%	71%
hubert-emotion-russian-resd	75%	92%
unispeech-sat-emotion-russian-resd Copied	72%	64%
wavlm-bert-base	81%	71%
wavlm-bert-fusion	83%	71%
Текстовые модели		
rubert-base-emotion-russian-cedr-m7	74%	43%
rubert-tiny2-russian-emotion-detection	85%	80%
rubert-large-emotion-russian-cedr-m7	76%	78%
rubert-tiny-emotion-russian-cedr-m7	72%	38%

Вывод: в ходе работы ознакомились с возможностями современных программных средств определения эмоционального фона речи человека в голосе и в тексте, получили навыки работы с открытой библиотекой искусственного интеллекта *Aniemore*, в результате проверки моделей библиотеки определили, что в модуле *Text* наивысшую точность показали модели архитектуры *Bert_Tiny2* (80%) и *Bert_Large* (78%), в модуле *Voice* — *Wav2Vec2* (99%) и *Hubert* (92%).