Tạo môi trường conda ảo để cố định phiên bản

```
# tải conda về và run
!wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
!bash Miniconda3-latest-Linux-x86_64.sh -bfp /usr/local

# cho phép conda tải về các gói
!/usr/local/bin/conda tos accept --override-channels --channel https://repo.anaconda.com/pkgs/main
!/usr/local/bin/conda tos accept --override-channels --channel https://repo.anaconda.com/pkgs/r

# khởi tạo môi trường conda
!/usr/local/bin/conda create -y -n py311 python=3.11

# demo chạy thử
!bash -c "source /usr/local/bin/activate py311 && python --version"
```

Hiện kết quả đã ẩn

Lấy repo mẫu và chỉnh sửa

```
!git clone https://github.com/hungithust/basic-multimodal-speech-emotion-recognition
```

Hiện kết quả đã ẩn

```
# cài đặt các thư viện cần thiết
%cd basic-multimodal-speech-emotion-recognition
!bash -c 'source /usr/local/bin/activate py311 && pip install -r requirements.txt'
```

Hiện kết quả đã ẩn

Thử nghiệm với bộ dữ liệu IEMOCAP

↳ 37 ô bị ẩn

Thực hiện với bộ dữ liệu tự làm

REPO này chạy hết bằng CPU

TODO:

1. Xử lý data rồi lưu vào thành audio_and_text.pkl
2. Tìm hiểu tại sao lại có frame khác nhau 91 - 170 dù data vẫn thế
3. Tìm xem các mô hình đầu ra là bao nhiêu để xử lý lớp linear phù hợp
4. Tìm cách chạy trên GPU (nếu được)

Mount data từ drive

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
# chuyển data sang thư mục data của repo
!mkdir /content/basic-multimodal-speech-emotion-recognition/data
```

```
!mkdir /content/basic-multimodal-speech-emotion-recognition/data/raw
!mkdir /content/basic-multimodal-speech-emotion-recognition/data/raw/Preprocessed
!cp -r /content/drive/MyDrive/project3/Preprocessed/* /content/basic-multimodal-speech-emotion-recognition/data/raw/Pr
```

```
!cp -r /content/basic-multimodal-speech-emotion-recognition/pictures/* /content/drive/MyDrive/project3/pictures
```

```
-r /content/basic-multimodal-speech-emotion-recognition/saved_models/fusion_state_dict.pt /content/drive/MyDrive/proj
```

```
# !rm -rf /content/basic-multimodal-speech-emotion-recognition/data/raw/happy+neutral+angry+sad
# !rm -rf /content/basic-multimodal-speech-emotion-recognition/data/raw/scare
# !rm -rf /content/basic-multimodal-speech-emotion-recognition/data/raw/surprise
```

⌄ Data processing

```
# import sys
# import os

# # Đường dẫn đến thư mục hiện tại (thư mục gốc của repo)
# current_dir = os.getcwd()

# # Thêm đường dẫn gốc vào sys.path
# if current_dir not in sys.path:
#     sys.path.append(current_dir)
#     print(f"Đã thêm đường dẫn gốc: {current_dir} vào sys.path")
```

```
Đã thêm đường dẫn gốc: /content/multimodal-speech-emotion-recognition vào sys.path
```

```
# tạo file setup.py để tìm và cài package, lưu ý phải để ở thư mục chính
!echo "from setuptools import setup, find_packages" > setup.py
!echo "" >> setup.py
!echo "setup(" >> setup.py
!echo "    name=\"mer\"," >> setup.py
!echo "    version=\"0.1.0\"," >> setup.py
!echo "    packages=find_packages()," >> setup.py
!echo ")" >> setup.py
```

```
# cài đặt các module thành package nhờ file setup.py
!bash -c 'source /usr/local/bin/activate py311 && pip install -e .'
```

```
Obtaining file:///content/basic-multimodal-speech-emotion-recognition
  Installing build dependencies ... done
  Checking if build backend supports build_editable ... done
  Getting requirements to build editable ... done
  Preparing editable metadata (pyproject.toml) ... done
Building wheels for collected packages: mer
  Building editable for mer (pyproject.toml) ... done
  Created wheel for mer: filename=mer-0.1.0-0.editable-py3-none-any.whl size=3573 sha256=ce898d3f7f22488ab0d87851e433
  Stored in directory: /tmp/pip-ephem-wheel-cache-zf13nd3p/wheels/e4/9c/df/84f23516cc1a1fb6d45b53f3cdb04149a9c581aaf6
Successfully built mer
Installing collected packages: mer
  Attempting uninstall: mer
    Found existing installation: mer 0.1.0
    Uninstalling mer-0.1.0:
      Successfully uninstalled mer-0.1.0
Successfully installed mer-0.1.0
```

Lưu ý: Chạy các file module thì đều có dạng:

!bash -c 'source py311/bin/activate && python -m mer.scripts.preprocess_data'

Cài đặt core

Chạy main theo từng bước

```
# # Đợi vài phút để PyPI cập nhật
# !pip install --upgrade protonx
```

```
# !bash -c 'source /usr/local/bin/activate py311 && pip install --upgrade protonx'
```

```
# bước 1 tạo file audio_and_text.pkl từ dataset
!bash -c 'source /usr/local/bin/activate py311 && python main.py'
```

```
# đổi tên file để backup
# !mv /content/multimodal-speech-emotion-recognition/data/IEMOCAP/Preprocessed/audio_and_text.pkl /content/multimoda
```

```
# df_moi = df.head(100)
# file_moi = '/content/multimodal-speech-emotion-recognition/data/IEMOCAP/Preprocessed/audio_and_text.pkl'
# df_moi.to_pickle(file_moi)
# print("đã lưu thành công")
```

```
đã lưu thành công
```

```
# bước 2
!bash -c 'source /usr/local/bin/activate py311 && python main.py'
```

```
/usr/local/envs/py311/lib/python3.11/site-packages/huggingface_hub/file_download.py:942: FutureWarning: `resume_downl
  warnings.warn(
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Turning audio into W2V2s:   0% 2/7274 [00:00<10:41, 11.34it/s]/usr/local/envs/py311/lib/python3.11/site-packages/hugg
  warnings.warn(
/usr/local/envs/py311/lib/python3.11/site-packages/huggingface_hub/file_download.py:942: FutureWarning: `resume_downl
  warnings.warn(
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Turning audio into W2V2s: 100% 7274/7274 [02:21<00:00, 51.37it/s]
```

Nhấp đúp (hoặc nhấn Enter) để chỉnh sửa

```
# bước 3
!bash -c 'source /usr/local/bin/activate py311 && python main.py'
```

```
/usr/local/envs/py311/lib/python3.11/site-packages/huggingface_hub/file_download.py:942: FutureWarning: `resume_downl
  warnings.warn(
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
vocab.txt: 895kB [00:00, 33.7MB/s]
bpe.codes: 1.14MB [00:00, 126MB/s]
tokenizer.json: 3.13MB [00:00, 184MB/s]
```

```
!bash -c 'source /usr/local/bin/activate py311 && pip install matplotlib-inline'
```

```
Collecting matplotlib-inline
  Downloading matplotlib_inline-0.2.1-py3-none-any.whl.metadata (2.3 kB)
Collecting traitlets (from matplotlib-inline)
  Downloading traitlets-5.14.3-py3-none-any.whl.metadata (10 kB)
Downloading matplotlib_inline-0.2.1-py3-none-any.whl (9.5 kB)
Downloading traitlets-5.14.3-py3-none-any.whl (85 kB)
Installing collected packages: traitlets, matplotlib-inline
  ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2/2 [matplotlib-inline]
Successfully installed matplotlib-inline-0.2.1 traitlets-5.14.3
```

```
# để đọc được parquet thì cần engine
!bash -c 'source /usr/local/bin/activate py311 && pip install pyarrow fastparquet'
```

Hiện kết quả đã ẩn

```
!bash -c 'source /usr/local/bin/activate py311 && pip install IPython'
```

Hiện kết quả đã ẩn

Lưu ý: Thêm để có thể chạy không GUI trên colab, và phải build lại package

import matplotlib

matplotlib.use("Agg")

Bắt đầu lập trình hoặc tạo mã bằng trí tuệ nhân tạo (AI).

```
# bước 4
!bash -c 'source /usr/local/bin/activate py311 && python main.py'
```

```
/usr/local/envs/py311/lib/python3.11/site-packages/huggingface_hub/file_download.py:942: FutureWarning: `resume_downl
  warnings.warn(
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Some weights of RobertaForSequenceClassification were not initialized from the model checkpoint at vinai/phobert-base
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
Đã khởi tạo xong model
Epoch 1:   0% 0/364 [00:00<?, ?it/s]We strongly recommend passing in an `attention_mask` since your input_ids may be
Epoch 1: 100% 364/364 [04:17<00:00,  1.41it/s, accuracy=0.273, loss=1.49]
Epoch 2: 100% 364/364 [04:22<00:00,  1.38it/s, accuracy=0.364, loss=1.42]
Epoch 3: 100% 364/364 [04:23<00:00,  1.38it/s, accuracy=0.364, loss=1.06]
Evaluating the model: 100% 91/91 [00:18<00:00,  4.84it/s]
Accuracy: 0.4824742268041237
```

Lưu ý: Fix linear thành 91 * 768, do frame của file audio khác nhau, gốc là 170, dữ liệu này là 91

```
# bước 5
!bash -c 'source /usr/local/bin/activate py311 && python main.py'
```

```
/usr/local/envs/py311/lib/python3.11/site-packages/huggingface_hub/file_download.py:942: FutureWarning: `resume_downl
  warnings.warn(
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
/usr/local/envs/py311/lib/python3.11/site-packages/huggingface_hub/file_download.py:942: FutureWarning: `resume_downl
  warnings.warn(
Some weights of Wav2Vec2Model were not initialized from the model checkpoint at facebook/wav2vec2-base-960h and are n
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
INFO:scripts.run_model:Getting the train dataloader...
INFO:dataloaders.iemocap:Loaded train dataset. Size: 5819
INFO:dataloaders.iemocap:Each emotion percentages: angry - 1182 | scare - 810 | neutral - 1190 | sad - 857 | surprise
INFO:dataloaders.iemocap:Shapes: audio - (39526,); text - (256,)
INFO:scripts.run_model:Starting the training process...
INFO:core.trainer:Training the audio model...
Epoch 1:   0% 0/364 [00:00<?, ?it/s]/usr/local/envs/py311/lib/python3.11/site-packages/torch/utils/data/_utils/collat
  return collate([torch.as_tensor(b) for b in batch], collate_fn_map=collate_fn_map)
Epoch 1: 100% 364/364 [02:21<00:00,  2.57it/s, accuracy=0.182, loss=1.86]
```

```
Epoch 2: 100% 364/364 [02:29<00:00,  2.43it/s, accuracy=0.636, loss=1.4]
Epoch 3: 100% 364/364 [02:29<00:00,  2.44it/s, accuracy=0.455, loss=1.58]
INFO:scripts.run_model:Saving the model...
INFO:scripts.run_model:Getting the eval dataloader...
INFO:dataloaders.iemocap:Loaded test dataset. Size: 1455
INFO:dataloaders.iemocap:Each emotion percentages: angry - 284 | scare - 184 | neutral - 317 | sad - 222 | surprise -
INFO:dataloaders.iemocap:Shapes: audio - (39526,); text - (256,)
INFO:scripts.run_model:Evaluating the model...
Evaluating the model: 100% 91/91 [00:15<00:00,  5.72it/s]
Accuracy: 0.5896907216494846
```

Lưu ý: Sửa fushion/model:

Phần mlp_head: sửa 170 thành 123 và bỏ đi +4

```
# bước 6
!bash -c 'source /usr/local/bin/activate py311 && python main.py'

/usr/local/envs/py311/lib/python3.11/site-packages/huggingface_hub/file_download.py:942: FutureWarning: `resume_downl
  warnings.warn(
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
Ignored unknown kwarg option normalize
/usr/local/envs/py311/lib/python3.11/site-packages/huggingface_hub/file_download.py:942: FutureWarning: `resume_downl
  warnings.warn(
Some weights of Wav2Vec2Model were not initialized from the model checkpoint at facebook/wav2vec2-base-960h and are n
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
INFO:scripts.run_model:Getting the train dataloader...
INFO:dataloaders.iemocap:Loaded train dataset. Size: 5819
INFO:dataloaders.iemocap:Each emotion percentages: angry - 1182 | scare - 810 | neutral - 1190 | sad - 857 | surprise -
INFO:dataloaders.iemocap:Shapes: audio - (39526,); text - (256,)
INFO:scripts.run_model:Starting the training process...
INFO:core.trainer:Training the fusion model...
Epoch 1:   0% 0/364 [00:00<?, ?it/s]/usr/local/envs/py311/lib/python3.11/site-packages/torch/utils/data/_utils/collat
  return collate([torch.as_tensor(b) for b in batch], collate_fn_map=collate_fn_map)
We strongly recommend passing in an `attention_mask` since your input_ids may be padded. See https://huggingface.co/d
Epoch 1: 100% 364/364 [02:27<00:00,  2.47it/s, accuracy=0.818, loss=1.34]
Epoch 2: 100% 364/364 [02:35<00:00,  2.34it/s, accuracy=0.727, loss=1.46]
Epoch 3: 100% 364/364 [02:35<00:00,  2.34it/s, accuracy=0.455, loss=1.65]
INFO:scripts.run_model:Saving the model...
INFO:scripts.run_model:Getting the eval dataloader...
INFO:dataloaders.iemocap:Loaded test dataset. Size: 1455
INFO:dataloaders.iemocap:Each emotion percentages: angry - 284 | scare - 184 | neutral - 317 | sad - 222 | surprise -
INFO:dataloaders.iemocap:Shapes: audio - (39526,); text - (256,)
INFO:scripts.run_model:Evaluating the model...
Evaluating the model: 100% 91/91 [00:35<00:00,  2.55it/s]
Accuracy: 0.5993127147766323
```

Bắt đầu lập trình hoặc tạo mã bằng trí tuệ nhân tạo (AI).

Bắt đầu lập trình hoặc tạo mã bằng trí tuệ nhân tạo (AI).

Bắt đầu lập trình hoặc tạo mã bằng trí tuệ nhân tạo (AI).

Bắt đầu lập trình hoặc tạo mã bằng trí tuệ nhân tạo (AI).

Bắt đầu lập trình hoặc tạo mã bằng trí tuệ nhân tạo (AI).

Bắt đầu lập trình hoặc tạo mã bằng trí tuệ nhân tạo (AI).

˅   Tạo dataframe rồi lưu vào dạng pkl trong data/preprocessed/audio_and_text.pkl

```
!bash -c 'source /usr/local/bin/activate py311 && python main.py'
```

Thử faster whisper và photonx

```
!bash -c 'source /usr/local/bin/activate py311 && pip install transformers accelerate datasets soundfile librosa &&
```

Hiện kết quả đã ẩn

```python
import torch
from transformers import WhisperProcessor, WhisperForConditionalGeneration
import librosa
import numpy as np

# 1. Định nghĩa tên mô hình PhoWhisper Small
model_name = "vinai/phowhisper-small"

# 2. Tải Processor và Model
# Processor sẽ tải tokenizer và feature extractor
processor = WhisperProcessor.from_pretrained(model_name)
model = WhisperForConditionalGeneration.from_pretrained(model_name)

# 3. Chuyển mô hình sang GPU nếu có
device = "cuda" if torch.cuda.is_available() else "cpu"
model.to(device)

print(f"Đã tải mô hình {model_name} và sử dụng trên thiết bị: {device}")
```

```
Đã tải mô hình vinai/phowhisper-small và sử dụng trên thiết bị: cpu
---------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
/tmp/ipython-input-4180741552.py in <cell line: 0>()
     58
     59 # transcribe_vietnamese_audio('female_00003.wav')
---> 60 segments,info = model.transcribe("female_00003.wav")
     61 for segment in segments:
     62     print("[%.2fs -> %.2fs] %s" % (segment.start, segment.end, segment.text))

/usr/local/lib/python3.12/dist-packages/torch/nn/modules/module.py in __getattr__(self, name)
   1962             if name in modules:
   1963                 return modules[name]
-> 1964         raise AttributeError(
   1965             f"'{type(self).__name__}' object has no attribute '{name}'"
   1966         )

AttributeError: 'WhisperForConditionalGeneration' object has no attribute 'transcribe'
```

```python
def transcribe_vietnamese_audio(audio_path):
    """
    Chuyển đổi tệp âm thanh Tiếng Việt sang văn bản bằng PhoWhisper Small.

    Args:
        audio_path (str): Đường dẫn đến tệp âm thanh (ví dụ: .wav, .mp3).

    Returns:
        str: Văn bản đã được chuyển đổi.
    """
    # 1. Tải và lấy mẫu lại (resample) tệp âm thanh về tần số 16kHz (yêu cầu của Whisper)
```

```python
    # librosa.load trả về numpy array và sample rate (sr)
    speech_array, sampling_rate = librosa.load(audio_path, sr=16000)

    # 2. Xử lý đầu vào: trích xuất features
    # Cần đảm bảo rằng audio là một mảng 1D (mono)
    input_features = processor(
        speech_array,
        sampling_rate=sampling_rate,
        return_tensors="pt"
    ).input_features

    # 3. Chuyển input features sang thiết bị (GPU/CPU)
    input_features = input_features.to(device)

    # 4. Chạy Inference (tạo văn bản)
    with torch.no_grad():
        # Tham số language="vi" và task="transcribe" là bắt buộc đối với PhoWhisper
        predicted_ids = model.generate(
            input_features,
            forced_decoder_ids=processor.get_decoder_prompt_ids(language="vi", task="transcribe")
        )

    # 5. Giải mã ID thành chuỗi văn bản
    transcription = processor.batch_decode(predicted_ids, skip_special_tokens=False)[0]

    return transcription


transcribe_vietnamese_audio('female_00003.wav')
```

'sau khi mà cho vào thì ông ấy đã đổ những viên đá nhỏ cùng với cát đặt bằng phẳng thì vừa vặn có thể đủ hết những cá
i viên đá vào bên trong.'

˅ thử emotion2vec

```python
!pip install funasr
from funasr import AutoModel
```

```
Collecting funasr
  Downloading funasr-1.2.7-py3-none-any.whl.metadata (32 kB)
Requirement already satisfied: scipy>=1.4.1 in /usr/local/lib/python3.12/dist-packages (from funasr) (1.16.3)
Requirement already satisfied: librosa in /usr/local/lib/python3.12/dist-packages (from funasr) (0.11.0)
Collecting jamo (from funasr)
  Downloading jamo-0.4.1-py3-none-any.whl.metadata (2.3 kB)
Requirement already satisfied: PyYAML>=5.1.2 in /usr/local/lib/python3.12/dist-packages (from funasr) (6.0.3)
Requirement already satisfied: soundfile>=0.12.1 in /usr/local/lib/python3.12/dist-packages (from funasr) (0.13.1)
Collecting kaldiio>=2.17.0 (from funasr)
  Downloading kaldiio-2.18.1-py3-none-any.whl.metadata (13 kB)
Collecting torch-complex (from funasr)
  Downloading torch_complex-0.4.4-py3-none-any.whl.metadata (3.1 kB)
Requirement already satisfied: sentencepiece in /usr/local/lib/python3.12/dist-packages (from funasr) (0.2.1)
Requirement already satisfied: jieba in /usr/local/lib/python3.12/dist-packages (from funasr) (0.42.1)
Collecting pytorch-wpe (from funasr)
  Downloading pytorch_wpe-0.0.1-py3-none-any.whl.metadata (242 bytes)
Requirement already satisfied: editdistance>=0.5.2 in /usr/local/lib/python3.12/dist-packages (from funasr) (0.8.1)
Collecting oss2 (from funasr)
  Downloading oss2-2.19.1.tar.gz (298 kB)
  ──────────────────────────────────────── 298.8/298.8 kB 7.9 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: tqdm in /usr/local/lib/python3.12/dist-packages (from funasr) (4.67.1)
Requirement already satisfied: umap-learn in /usr/local/lib/python3.12/dist-packages (from funasr) (0.5.9.post2)
Collecting jaconv (from funasr)
  Downloading jaconv-0.4.1-py3-none-any.whl.metadata (8.0 kB)
Collecting hydra-core>=1.3.2 (from funasr)
  Downloading hydra_core-1.3.2-py3-none-any.whl.metadata (5.5 kB)
Collecting tensorboardX (from funasr)
  Downloading tensorboardx-2.6.4-py3-none-any.whl.metadata (6.2 kB)
Requirement already satisfied: requests in /usr/local/lib/python3.12/dist-packages (from funasr) (2.32.4)
Collecting modelscope (from funasr)
  Downloading modelscope-1.33.0-py3-none-any.whl.metadata (43 kB)
  ──────────────────────────────────────── 43.3/43.3 kB 3.0 MB/s eta 0:00:00
Requirement already satisfied: omegaconf<2.4,>=2.2 in /usr/local/lib/python3.12/dist-packages (from hydra-core>=1.3.2
Requirement already satisfied: antlr4-python3-runtime==4.9.* in /usr/local/lib/python3.12/dist-packages (from hydra-c
Requirement already satisfied: packaging in /usr/local/lib/python3.12/dist-packages (from hydra-core>=1.3.2->funasr)
Requirement already satisfied: numpy in /usr/local/lib/python3.12/dist-packages (from kaldiio>=2.17.0->funasr) (2.0.2
Requirement already satisfied: cffi>=1.0 in /usr/local/lib/python3.12/dist-packages (from soundfile>=0.12.1->funasr)
Requirement already satisfied: audioread>=2.1.9 in /usr/local/lib/python3.12/dist-packages (from librosa->funasr) (3.
Requirement already satisfied: numba>=0.51.0 in /usr/local/lib/python3.12/dist-packages (from librosa->funasr) (0.60.
Requirement already satisfied: scikit-learn>=1.1.0 in /usr/local/lib/python3.12/dist-packages (from librosa->funasr)
Requirement already satisfied: joblib>=1.0 in /usr/local/lib/python3.12/dist-packages (from librosa->funasr) (1.5.2)
Requirement already satisfied: decorator>=4.3.0 in /usr/local/lib/python3.12/dist-packages (from librosa->funasr) (4.
Requirement already satisfied: pooch>=1.1 in /usr/local/lib/python3.12/dist-packages (from librosa->funasr) (1.8.2)
Requirement already satisfied: soxr>=0.3.2 in /usr/local/lib/python3.12/dist-packages (from librosa->funasr) (1.0.0)
Requirement already satisfied: typing_extensions>=4.1.1 in /usr/local/lib/python3.12/dist-packages (from librosa->fun
Requirement already satisfied: lazy_loader>=0.1 in /usr/local/lib/python3.12/dist-packages (from librosa->funasr) (0.
Requirement already satisfied: msgpack>=1.0 in /usr/local/lib/python3.12/dist-packages (from librosa->funasr) (1.1.2)
Requirement already satisfied: filelock in /usr/local/lib/python3.12/dist-packages (from modelscope->funasr) (3.20.0)
Requirement already satisfied: setuptools in /usr/local/lib/python3.12/dist-packages (from modelscope->funasr) (75.2.
Requirement already satisfied: urllib3>=1.26 in /usr/local/lib/python3.12/dist-packages (from modelscope->funasr) (2.
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests->fu
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests->funasr) (3.11)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from requests->funasr)
Collecting crcmod>=1.7 (from oss2->funasr)
  Downloading crcmod-1.7.tar.gz (89 kB)
  ──────────────────────────────────────── 89.7/89.7 kB 6.8 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Collecting pycryptodome>=3.4.7 (from oss2->funasr)
  Downloading pycryptodome-3.23.0-cp37-abi3-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (3.4 kB)
Collecting aliyun-python-sdk-kms>=2.4.1 (from oss2->funasr)
  Downloading aliyun_python_sdk_kms-2.16.5-py2.py3-none-any.whl.metadata (1.5 kB)
Collecting aliyun-python-sdk-core>=2.13.12 (from oss2->funasr)
  Downloading aliyun-python-sdk-core-2.16.0.tar.gz (449 kB)
  ──────────────────────────────────────── 449.6/449.6 kB 19.0 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: six in /usr/local/lib/python3.12/dist-packages (from oss2->funasr) (1.17.0)
Requirement already satisfied: protobuf>=3.20 in /usr/local/lib/python3.12/dist-packages (from tensorboardX->funasr)
Requirement already satisfied: pynndescent>=0.5 in /usr/local/lib/python3.12/dist-packages (from umap-learn->funasr)
Collecting jmespath<1.0.0,>=0.9.3 (from aliyun-python-sdk-core>=2.13.12->oss2->funasr)
  Downloading jmespath-0.10.0-py2.py3-none-any.whl.metadata (8.0 kB)
Requirement already satisfied: cryptography>=3.0.0 in /usr/local/lib/python3.12/dist-packages (from aliyun-python-sdk
Requirement already satisfied: pycparser in /usr/local/lib/python3.12/dist-packages (from cffi>=1.0->soundfile>=0.12.
Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in /usr/local/lib/python3.12/dist-packages (from numba>=0.5
Requirement already satisfied: platformdirs>=2.5.0 in /usr/local/lib/python3.12/dist-packages (from pooch>=1.1->libro
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.12/dist-packages (from scikit-learn>=1.
Downloading funasr-1.2.7-py3-none-any.whl (703 kB)
  ──────────────────────────────────────── 703.2/703.2 kB 41.7 MB/s eta 0:00:00
```

```
Downloading hydra_core-1.3.2-py3-none-any.whl (154 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 154.5/154.5 kB 11.9 MB/s eta 0:00:00
Downloading kaldiio-2.18.1-py3-none-any.whl (29 kB)
Downloading jaconv-0.4.1-py3-none-any.whl (15 kB)
Downloading jamo-0.4.1-py3-none-any.whl (9.5 kB)
Downloading modelscope-1.33.0-py3-none-any.whl (6.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6.1/6.1 MB 113.2 MB/s eta 0:00:00
Downloading pytorch_wpe-0.0.1-py3-none-any.whl (8.1 kB)
Downloading tensorboardx-2.6.4-py3-none-any.whl (87 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 87.2/87.2 kB 6.2 MB/s eta 0:00:00
Downloading torch_complex-0.4.4-py3-none-any.whl (9.1 kB)
Downloading aliyun_python_sdk_kms-2.16.5-py2.py3-none-any.whl (99 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 99.5/99.5 kB 7.9 MB/s eta 0:00:00
Downloading pycryptodome-3.23.0-cp37-abi3-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (2.3 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.3/2.3 MB 55.0 MB/s eta 0:00:00
Downloading jmespath-0.10.0-py2.py3-none-any.whl (24 kB)
Building wheels for collected packages: oss2, aliyun-python-sdk-core, crcmod
  Building wheel for oss2 (setup.py) ... done
  Created wheel for oss2: filename=oss2-2.19.1-py3-none-any.whl size=123940 sha256=1b6177c0e5a99e0cfa1bc1c2b1c11f42c3
  Stored in directory: /root/.cache/pip/wheels/a6/aa/ce/a0c9e73f8e4a3b7813b6b0d9dbddfb83028fd416828f27d97b
  Building wheel for aliyun-python-sdk-core (setup.py) ... done
  Created wheel for aliyun-python-sdk-core: filename=aliyun_python_sdk_core-2.16.0-py3-none-any.whl size=535315 sha25
  Stored in directory: /root/.cache/pip/wheels/5a/60/7c/80d5fdcd6d0e016e4a7ff4a66fd9321b3096ae676d78fe0212
  Building wheel for crcmod (setup.py) ... done
  Created wheel for crcmod: filename=crcmod-1.7-cp312-cp312-linux_x86_64.whl size=31834 sha256=35980badbe8f16d061cf3e
  Stored in directory: /root/.cache/pip/wheels/76/08/0b/caa8b1380122cbfe6a03eaccbec0f63c67e619af4e30ca5e2a
Successfully built oss2 aliyun-python-sdk-core crcmod
Installing collected packages: jamo, jaconv, crcmod, torch-complex, tensorboardX, pytorch-wpe, pycryptodome, kaldiio,
Successfully installed aliyun-python-sdk-core-2.16.0 aliyun-python-sdk-kms-2.16.5 crcmod-1.7 funasr-1.2.7 hydra-core-
/usr/local/lib/python3.12/dist-packages/pydub/utils.py:300: SyntaxWarning: invalid escape sequence '\('
  m = re.match('([su]([0-9]{1,2})p?) \(([0-9]{1,2}) bit\)$', token)
/usr/local/lib/python3.12/dist-packages/pydub/utils.py:301: SyntaxWarning: invalid escape sequence '\('
  m2 = re.match('([su]([0-9]{1,2})p?)( \(default\))?$', token)
/usr/local/lib/python3.12/dist-packages/pydub/utils.py:310: SyntaxWarning: invalid escape sequence '\('
  elif re.match('(flt)p?( \(default\))?$', token):
/usr/local/lib/python3.12/dist-packages/pydub/utils.py:314: SyntaxWarning: invalid escape sequence '\('
  elif re.match('(dbl)p?( \(default\))?$', token):
funasr version: 1.2.7.
Check update of funasr, and it would cost few times. You may disable it by set `disable_update=True` in AutoModel
You are using the latest version of funasr-1.2.7
Downloading Model from https://www.modelscope.cn to directory: /root/.cache/modelscope/hub/models/iic/emotion2vec_plu
2025-12-14 07:24:40,859 - modelscope - INFO - Got 11 files, start to download ...
Processing 11 items:  91%                                    10.0/11.0 [04:42<00:00, 1.67it/s]

Downloading [config.yaml]: 100%                              5.42k/5.42k [00:01<00:00, 5.29kB/s]

Downloading [.DS_Store]: 100%                               6.00k/6.00k [00:01<00:00, 4.46kB/s]

Downloading [configuration.json]: 100%                       343/343 [00:01<00:00, 248B/s]

Downloading [emotion2vec+data.png]: 100%                     267k/267k [00:01<00:00, 191kB/s]

Downloading [emotion2vec+radar.png]: 100%                    666k/666k [00:01<00:00, 513kB/s]

Downloading [logo.png]: 100%                                 1.76M/1.76M [00:01<00:00, 654kB/s]

Downloading [model.pt]: 100%                                 1.81G/1.81G [04:41<00:00, 5.80MB/s]

Downloading [README.md]: 100%                                5.26k/5.26k [00:01<00:00, 4.08kB/s]

Downloading [requirements.txt]: 100%                         43.0/43.0 [00:03<00:00, 13.5B/s]

Downloading [example/test.wav]: 100%                         128k/128k [00:01<00:00, 69.7kB/s]

Downloading [tokens.txt]: 100%                               119/119 [00:01<00:00, 87.9B/s]
---------------------------------------------------------------------------
KeyboardInterrupt                         Traceback (most recent call last)
/tmp/ipython-input-2893262403.py in <cell line: 0>()
      2 from funasr import AutoModel
      3
----> 4 model = AutoModel(model="iic/emotion2vec_plus_large")
      5
      6 wav_file = f"{model.model_path}/example/test.wav"

                    ⌃⌄ 11 frames
/usr/lib/python3.12/threading.py in wait(self, timeout)
    353         try:    # restore state no matter what (e.g., KeyboardInterrupt)
    354             if timeout is None:
--> 355                 waiter.acquire()
```

```python
from funasr import AutoModel

model = AutoModel(model="iic/emotion2vec_plus_base")

wav_file = "female_00003.wav"
res = model.generate(wav_file, output_dir="./outputs", granularity="utterance", extract_embedding=True)
print(res)
```

```
funasr version: 1.2.7.
Check update of funasr, and it would cost few times. You may disable it by set `disable_update=True` in AutoModel
You are using the latest version of funasr-1.2.7
Downloading Model from https://www.modelscope.cn to directory: /root/.cache/modelscope/hub/models/iic/emotion2vec_plu
2025-12-14 07:37:21,580 - modelscope - INFO - Got 10 files, start to download ...

Processing 10 items: 100%                                         10.0/10.0 [01:42<00:00, 28.8s/it]

Downloading [config.yaml]: 100%                                   3.10k/3.10k [00:01<00:00, 2.75kB/s]

Downloading [emotion2vec+data.png]: 100%                          267k/267k [00:01<00:00, 181kB/s]

Downloading [.DS_Store]: 100%                                     6.00k/6.00k [00:01<00:00, 4.58kB/s]

Downloading [emotion2vec+radar.png]: 100%                         666k/666k [00:01<00:00, 391kB/s]

Downloading [logo.png]: 100%                                      1.76M/1.76M [00:01<00:00, 755kB/s]

Downloading [configuration.json]: 100%                            343/343 [00:01<00:00, 316B/s]

Downloading [model.pt]: 100%                                      1.04G/1.04G [01:41<00:00, 10.1MB/s]

Downloading [README.md]: 100%                                     5.26k/5.26k [00:00<00:00, 5.88kB/s]

Downloading [example/test.wav]: 100%                              128k/128k [00:01<00:00, 74.8kB/s]

Downloading [tokens.txt]: 100%                                    120/120 [00:00<00:00, 137B/s]

2025-12-14 07:39:04,394 - modelscope - INFO - Download model 'iic/emotion2vec_plus_base' successfully.
WARNING:root:trust_remote_code: False
Warning, miss key in ckpt: modality_encoders.AUDIO.decoder.blocks.0.0.weight, /root/.cache/modelscope/hub/models/iic/
Warning, miss key in ckpt: modality_encoders.AUDIO.decoder.blocks.0.0.bias, /root/.cache/modelscope/hub/models/iic/em
Warning, miss key in ckpt: modality_encoders.AUDIO.decoder.blocks.1.0.weight, /root/.cache/modelscope/hub/models/iic/
Warning, miss key in ckpt: modality_encoders.AUDIO.decoder.blocks.1.0.bias, /root/.cache/modelscope/hub/models/iic/em
Warning, miss key in ckpt: modality_encoders.AUDIO.decoder.blocks.2.0.weight, /root/.cache/modelscope/hub/models/iic/
Warning, miss key in ckpt: modality_encoders.AUDIO.decoder.blocks.2.0.bias, /root/.cache/modelscope/hub/models/iic/em
Warning, miss key in ckpt: modality_encoders.AUDIO.decoder.blocks.3.0.weight, /root/.cache/modelscope/hub/models/iic/
Warning, miss key in ckpt: modality_encoders.AUDIO.decoder.blocks.3.0.bias, /root/.cache/modelscope/hub/models/iic/em
Warning, miss key in ckpt: modality_encoders.AUDIO.decoder.proj.weight, /root/.cache/modelscope/hub/models/iic/emotio
Warning, miss key in ckpt: modality_encoders.AUDIO.decoder.proj.bias, /root/.cache/modelscope/hub/models/iic/emotion2
rtf_avg: 0.309: 100%|████████████| 1/1 [00:02<00:00, 2.29s/it][{'key': 'female_00003', 'labels': ['生气/angry', '厌恶/d
         2.43408591e-01,  2.91353106e-01, -1.58451393e-01,  1.63543552e-01,
         3.20388526e-01, -5.21452986e-02, -1.24764752e+00,  1.68536520e+00,
        -1.37340176e+00, -3.32276678e+00,  1.69772637e+00,  8.76020193e-02,
         6.48822844e-01, -1.17778718e-01,  1.82752192e+00, -3.64174008e-01,
         1.15052819e+00, -5.12462735e-01,  2.53382969e+00,  1.16178858e+00,
        -1.82023251e+00, -3.86909038e-01,  1.52495396e+00,  7.93120742e-01,
        -1.55946624e+00, -1.05707312e+00, -2.09639478e+00, -2.30493665e-01,
        -5.78809142e-01,  5.90470918e-02,  1.93253994e+00,  1.02531052e+00,
        -1.05847263e+00, -2.12996674e+00, -4.03982908e-01, -1.51834893e+00,
         2.80221772e+00,  4.55187738e-01, -2.57587433e+00, -6.85958028e-01,
         5.47775209e-01, -1.07142246e+00, -1.17795801e+00, -1.82989156e+00,
        -3.09981674e-01,  1.45845008e+00,  2.32178420e-01,  8.86423945e-01,
         7.52838969e-01,  2.80687124e-01, -7.50850797e-01,  1.13865113e+00,
        -8.11983943e-01, -1.29478371e+00,  7.86235154e-01,  1.87536255e-01,
         1.16540384e+00, -1.28875220e+00,  5.75723171e-01, -7.86357895e-02,
         9.84320402e-01,  6.55718088e-01,  9.65957642e-01,  8.75313699e-01,
         2.53554340e-02, -1.05517045e-01, -1.70441973e+00,  1.10329878e+00,
        -5.25261685e-02,  1.68407822e+00, -2.21487656e-01,  9.39819694e-01,
        -7.77691424e-01, -2.16068769e+00, -1.45558357e+00,  1.85968280e+00,
        -5.69783986e-01, -1.58319616e+00,  4.23533499e-01,  1.18876624e+00,
         9.16862667e-01,  6.59008145e-01,  2.31853202e-01,  1.63366199e-02,
         4.53197986e-01, -5.47773719e-01, -1.12742019e+00, -1.75995517e+00,
        -1.25376940e+00,  1.78288853e+00, -1.17143261e+00,  3.82506967e-01,
         8.39173913e-01,  2.13485226e-01, -1.91783905e+00,  1.51709187e+00,
         1.59195554e+00, -1.31531417e+00, -8.48372057e-02,  3.09348881e-01,
        -3.86421293e-01, -1.61757231e+00, -1.38614023e+00,  3.49597365e-01,
        -7.43109107e-01,  9.26456690e-01,  4.94756341e-01,  1.54274428e+00,
        -1.15518904e+00, -1.13341331e+00,  7.57633507e-01, -5.39519548e-01,
         9.54601943e-01, -1.61045933e+00, -6.24073565e-01, -8.21466982e-01,
        -1.68676209e+00, -3.46004218e-01, -8.90741765e-01,  2.02841353e+00,
        -9.67915714e-01,  2.48645544e-02,  2.15644464e-01, -1.35437584e+00,
        -8.34677041e-01, -1.91323888e+00, -1.07356763e+00, -6.87668443e-01,
        -9.08678114e-01, -3.72025371e-01, -1.06177545e+00,  9.25996542e-01,
         1.79084146e+00,  6.35644019e-01, -4.65113930e-02, -1.03331304e+00,
         6.18319511e-01, -6.17523909e-01, -2.25875169e-01,  2.54504643e-02,
         1.06574357e+00, -5.18130839e-01, -1.61688849e-01,  1.19866729e+00,
        -1.34309381e-01, -1.11932421e+00,  1.33714452e-01,  4.99570817e-01,
        -6.37145638e-01, -2.25058651e+00, -1.36659393e-04, -1.22479451e+00,
        -1.03847456e+00,  1.37129819e+00,  1.25816166e+00, -1.52755558e-01,
        -1.70541644e+00,  4.05562371e-01,  1.74303341e+00,  1.30491912e+00,
```

```
 8.56494546e-01,  1.24784267e+00, -1.42014885e+00,  3.09445143e-01,
-5.26000500e-01,  6.11652434e-01, -3.33926268e-02, -3.29129435e-02,
-7.79197991e-01,  2.07210946e+00,  6.87312007e-01,  4.48227704e-01,
-6.42678380e-01, -7.23094523e-01, -4.11716074e-01, -3.98215145e-01,
-3.03348899e-01, -1.36557385e-01, -1.08609796e+00,  9.54325020e-01,
 1.31883085e+00, -6.99394107e-01,  5.23525357e-01, -1.11177742e-01,
 4.63655144e-01,  1.51412821e+00, -3.59259993e-02,  4.30253357e-01,
 4.57130484e-02, -4.44040298e+00,  8.74838650e-01, -5.07884741e-01,
 4.03180748e-01, -1.36955369e+00, -1.56954706e+00, -5.08861959e-01,
-1.49242985e+00,  2.72974730e-01, -5.39476946e-02,  7.98983693e-01,
 2.06797862e+00, -1.98637620e-01,  1.81973088e+00,  1.26926744e+00,
-7.35904574e-01, -1.31459200e+00,  4.05526042e-01, -1.45292848e-01,
-2.36899942e-01,  1.04461873e+00, -2.74907589e-01, -1.20994902e+00,
-2.54219919e-01,  9.43183184e-01,  4.77194500e+00,  5.94889939e-01,
-4.38395977e-01,  2.08385372e+00, -6.44905448e-01,  7.21467733e-01,
-9.22652721e-01,  9.83312964e-01, -1.56669378e+00,  1.39451742e+00,
 2.04388857e+00,  2.58085430e-01, -1.00462079e+00, -4.75616157e-01,
 1.20797503e+00,  4.23860133e-01,  1.55830371e+00,  1.31303763e+00,
 1.45371938e+00, -4.68820482e-01, -7.45138526e-01,  8.25827360e-01,
 1.02689946e+00,  7.13733971e-01,  1.30890533e-01, -6.77771926e-01,
-4.56089705e-01, -1.31367731e+00,  1.72973000e+00,  1.37420905e+00,
-1.65748620e+00,  3.26552927e-01, -8.83334205e-02,  1.98437721e-02,
-1.43356669e+00,  6.10122323e-01,  1.61742818e+00,  1.90055883e+00,
 1.43562841e+00,  1.89976597e+00,  1.81157935e+00, -6.40498459e-01,
 1.14204001e+00, -8.74229297e-02,  9.85445976e-01,  1.24533439e+00,
 1.83497250e-01, -1.09633875e+00, -1.04929650e+00, -3.65905672e-01,
 3.59793217e-03, -6.55249536e-01,  6.29109502e-01, -4.05539930e-01,
-5.88036835e-01,  2.06554737e-02, -3.16540033e-01,  1.05654144e+00,
-1.85060966e+00,  1.53399801e+00,  1.11645949e+00, -3.00930291e-01,
 2.55166680e-01, -1.80552614e+00,  1.95897031e+00, -5.64876318e-01,
-1.49526238e+00, -6.08184077e-02,  1.43559575e+00, -2.03315520e+00,
-8.20595801e-01, -6.77640617e-01,  1.12356591e+00, -9.78011608e-01,
 3.77976441e+00, -6.62492335e-01,  1.03313470e+00, -4.24227089e-01,
 6.13133729e-01, -1.53999555e+00,  6.29241616e-02,  1.09124970e+00,
-1.40906417e+00,  1.85580397e+00, -7.49807000e-01,  5.17808676e-01,
-9.57105041e-01,  8.96697998e-01, -1.15626955e+00,  8.03909481e-01,
-1.01928389e+00,  1.59597754e-01, -5.26271701e-01,  3.91310990e-01,
 5.32803059e-01, -2.25367080e-02, -5.53814173e-01, -1.17498338e+00,
-4.07690816e-02,  1.31259406e+00,  7.81728685e-01, -6.32618487e-01,
 1.32244790e+00, -7.60487378e-01,  7.89868474e-01,  1.09910226e+00,
 4.08142954e-01,  1.54978323e+00,  1.35140657e+00, -1.37368739e-01,
-7.26731360e-01, -1.57292640e+00,  1.05531231e-01,  1.27245438e+00,
-5.01565874e-01, -4.97380018e-01,  1.17113686e+00, -1.56168473e+00,
 1.78500748e+00, -1.14445484e+00, -3.38209301e-01, -6.07603312e-01,
 2.78451070e-02, -3.30807298e-01, -1.14724946e+00, -6.53911650e-01,
 1.07833371e-01,  7.46341527e-01, -1.04031003e+00, -1.27687871e+00,
-3.35225940e-01, -4.20343727e-01,  8.45062733e-01,  1.86897182e+00,
-6.33305430e-01,  5.39660096e-01, -3.33845645e-01,  1.66927958e+00,
-2.91186452e-01, -5.12776971e-01, -2.37890378e-01, -1.12070404e-01,
 7.52344072e-01,  4.76840198e-01,  3.31067681e-01,  5.19596159e-01,
 9.14266884e-01,  5.12291312e-01, -1.45530212e+00,  1.38649857e+00,
-1.45650983e+00,  5.72996795e-01,  5.89272022e-01,  8.55897889e-02,
 5.59137225e-01,  8.81235957e-01, -1.20812953e+00, -4.12039429e-01,
-4.57383811e-01, -1.13369370e+00,  3.30063969e-01, -8.55427742e-01,
 2.68814415e-01, -9.68080938e-01, -2.47407153e-01, -3.51619482e-01,
-1.26032066e+00,  6.96264148e-01,  1.90545142e+00, -1.98844910e-01,
 7.50802040e-01,  1.57959127e+00, -7.30015710e-01, -2.25218940e+00,
-1.60916102e+00, -2.09535480e+00,  1.79807210e+00,  1.41943312e+00,
 4.30133730e-01,  1.15220046e+00,  2.25776568e-01,  1.87539685e+00,
 1.10482872e+00,  1.69701111e+00, -1.75462317e+00,  4.50065613e-01,
 6.49527133e-01,  6.06437624e-01, -1.65574396e+00,  7.37051368e-01,
 1.06020057e+00,  1.42577338e+00,  6.17114842e-01,  2.87384212e-01,
-1.43619895e+00,  2.00700145e-02,  1.26150894e+00,  2.22733021e+00,
 1.39951336e+00, -3.32530141e-01,  3.39041129e-02, -1.82457387e+00,
 6.20194912e-01,  3.71798128e-01, -1.58926094e+00, -1.21101058e+00,
 3.62187475e-02,  1.86174560e+00,  9.90411714e-02, -9.09975111e-01,
 4.34599183e-02,  5.11253297e-01,  1.47864377e+00,  1.39080310e+00,
-2.12632373e-01, -1.29532707e+00,  4.23366934e-01,  9.50547695e-01,
 5.50185025e-01,  6.49099410e-01,  1.04724860e+00,  6.98433518e-01,
-1.19451320e+00,  1.12547599e-01,  1.55471277e+00,  1.74839962e+00,
 1.44498120e-03, -8.17997977e-02, -1.27904987e+00,  1.25484073e+00,
 2.70550609e-01,  5.25582694e-02,  8.17855895e-01,  1.07565176e+00,
 1.67415962e-02, -5.29345274e-01,  8.63631368e-01, -5.58948517e-01,
 6.32795632e-01,  2.74146020e-01,  8.12688112e-01,  7.19642043e-02,
-9.70925912e-02,  1.76962122e-01,  6.40526637e-02,  1.04152203e+00,
-5.22654235e-01,  4.58174258e-01, -2.56529242e-01, -3.11031103e-01,
 5.20243287e-01, -3.28779733e-03,  2.32221842e-01,  1.85299778e+00,
 1.63955224e+00, -6.13707423e-01,  3.26257288e-01, -2.05533409e+00,
```

```
        -1.21822441e+00,  6.51529135e-02,  1.24357930e+00,  3.15027728e-01,
```

```
print(len(res[0]['feats']))
```

```
768     -1.24701858e+00, -2.65036966e-01, -8.71381164e-01,  2.31540203e-01,
        -6.88744485e-01, -2.64537066e-01, -1.18945158e+00,  6.64517358e-02,
         1.59207714e+00,  6.97901368e-01,  1.02416420e+00,  3.80037040e-01,
```

```python
from transformers import Wav2Vec2ForPreTraining

# Tải mô hình Pre-trained (ví dụ: bản tiếng Anh/đa ngôn ngữ)
# Chúng ta không cần AutoModelForCTC vì không muốn tạo văn bản
model_id_feature = "facebook/wav2vec2-base"

processor_feature = AutoProcessor.from_pretrained(model_id_feature)
model_feature = Wav2Vec2ForPreTraining.from_pretrained(model_id_feature)
device = "cuda" if torch.cuda.is_available() else "cpu"
model_feature.to(device)

# def load_and_preprocess_audio(audio_path: str, target_sr: int = 16000) -> np.ndarray:
#     """Tải tệp âm thanh và lấy mẫu lại về tần số 16kHz."""
#     speech_array, sr = librosa.load(audio_path, sr=None)

#     # Lấy mẫu lại nếu cần
#     # if sr != target_sr:
#     #     speech_array = librosa.resample(speech_array, orig_sr=sr, target_sr=target_sr)

#     return speech_array

# Thay bằng đường dẫn đến tệp WAV/MP3 của bạn
sample_audio_path = 'female_00003.wav'
speech_array, sr = librosa.load(sample_audio_path, sr=16000)
input_audio = speech_array

# 1. Xử lý đầu vào
inputs_feature = processor_feature(
    input_audio,
    sampling_rate=16000,
    return_tensors="pt",
    padding=True
)

inputs_feature = {key: val.to(device) for key, val in inputs_feature.items()}

# 2. Chạy mô hình để lấy đầu ra (output)
with torch.no_grad():
    # output_hidden_states=True là quan trọng để lấy các layers ẩn
    outputs_feature = model_feature(**inputs_feature, output_hidden_states=True)

# 3. Lấy Đặc trưng (Features/Embeddings)
# Lấy hidden states từ lớp cuối cùng
last_hidden_state = outputs_feature.hidden_states[-1]

# last_hidden_state có shape: (batch_size, sequence_length, hidden_size)
# Ví dụ: (1, 199, 768) - nghĩa là 199 vector, mỗi vector 768 chiều
print("\n--- KẾT QUẢ TRÍCH XUẤT ĐẶC TRƯNG ---")
print(f"Shape của Last Hidden State: {last_hidden_state.shape}")

# Nếu bạn muốn lấy vector đặc trưng trung bình cho toàn bộ âm thanh:
audio_embedding = last_hidden_state.mean(dim=1).squeeze().cpu().numpy()
print(f"Shape của Audio Embedding (Vector trung bình): {audio_embedding.shape}")
```

```
        4.68055367e-01, -9.65257287e-01,  3.12693477e-01,  1.73984122e+00,
       -1.34550297e+00, -7.05765426e-01, -1.47963655e+00,  3.64977449e-01,
--- KẾT QUẢ TRÍCH XUẤT ĐẶC TRƯNG ---
        4.69980299e-01, -1.24316263e+00, -9.39774692e-01, -1.31460500e+00,
Shape của Last Hidden State: torch.Size([1, 379, 768])
        1.79215770e+00,  1.11299827e+00, -6.55721604e-01,  7.52043426e-02,
Shape của Audio Embedding (Vector trung bình): (768,)
       -1.08482266e-01, -1.86567079e+00, -1.45827603e-01, -1.27698588e+00,
       -1.60652387e+00,  5.28925717e-01, -5.04601419e-01, -6.95137620e-01],
```

```python
from funasr import AutoModel
import torch
import librosa
```

```
DEVICE = "cuda" if torch.cuda.is_available() else "cpu"

# ===== Load model =====
model = AutoModel(
    model="iic/emotion2vec_plus_base"
)

model.model.eval()

# ===== Load audio =====
speech, sr = librosa.load("sohai2H.wav", sr=16000)
speech_tensor = torch.from_numpy(speech).float().unsqueeze(0).to(DEVICE)

# ===== Build masks =====
padding_mask = torch.zeros(
    speech_tensor.shape,
    dtype=torch.bool,
    device=speech_tensor.device
)

# ===== Audio encoder =====
audio_encoder = model.model.modality_encoders["AUDIO"]

with torch.no_grad():
    # 1. Audio encoder

    x = audio_encoder(
        speech_tensor,
        padding_mask=padding_mask,
        mask=None,
        remove_masked=False
    )['x']


    # 2. Dropout input
    x = model.model.dropout_input(x)

    print(type(x))

    # 3. Transformer backbone
    for i, blk in enumerate(model.model.blocks):
        print(f'lan {i}: {type(x)}: {x}')
        x = blk(x)[0]

# ===== Last hidden state =====
last_hidden_state = x
print("Last hidden state shape:", last_hidden_state.shape)
print(last_hidden_state.mean(dim =1))
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
/tmp/ipython-input-999107379.py in <cell line: 0>()
      6
      7 # ===== Load model =====
----> 8 model = AutoModel.from_pretrained(
      9     model="iic/emotion2vec_plus_base",
     10     mask_time_first=True,

AttributeError: type object 'AutoModel' has no attribute 'from_pretrained'
```

```
print(model.model)
```

```
          (drop2): Dropout(p=0.0, inplace=False)
        )
        (post_mlp_dropout): Dropout(p=0.1, inplace=False)
      )
    )
    (norm): LayerNorm((768,), eps=1e-05, elementwise_affine=True)
    (dropout): Dropout(p=0.1, inplace=True)
  )
  (decoder): Decoder1d(
    (blocks): Sequential(
      (0): Sequential(
        (0): Conv1d(768, 384, kernel_size=(7,), stride=(1,), padding=(3,), groups=16)
        (1): SamePad()
        (2): TransposeLast()
        (3): LayerNorm((384,), eps=1e-05, elementwise_affine=False)
        (4): TransposeLast()
        (5): GELU(approximate='none')
      )
      (1): Sequential(
        (0): Conv1d(384, 384, kernel_size=(7,), stride=(1,), padding=(3,), groups=16)
        (1): SamePad()
        (2): TransposeLast()
        (3): LayerNorm((384,), eps=1e-05, elementwise_affine=False)
        (4): TransposeLast()
```