

Xử lý dữ liệu âm thanh

Hướng dẫn này hiển thị các phương pháp cụ thể để xử lý tập dữ liệu âm thanh. Tìm hiểu cách:

- Lấy mẫu lại tốc độ lấy mẫu.
- Use `map()` with audio datasets.

Để biết hướng dẫn về cách xử lý bất kỳ loại tập dữ liệu nào, hãy xem hướng dẫn quy trình chung.

Dàn diễn viên

The `cast_column()` function is used to cast a column to another feature to be decoded. When bạn sử dụng chức năng này với tính năng Âm thanh, bạn có thể lấy mẫu lại tốc độ lấy mẫu:

```
>>> from datasets import load_dataset, Audio

>>> dataset = load_dataset("PolyAI/minds14", "en-US", split="train")
>>> dataset = dataset.cast_column("audio", Audio(sampling_rate=16000))
```

Các tệp âm thanh được giải mã và lấy mẫu lại nhanh chóng, vì vậy lần tiếp theo bạn truy cập vào một ví dụ, tệp âm thanh được lấy mẫu lại thành 16kHz:

```
>>> audio = dataset[0]["audio"]
<datasets.features._torchcodec.AudioDecoder object at 0x11642b6a0>
>>> audio = audio_dataset[0]["audio"]
>>> samples = audio.get_all_samples()
>>> samples.data
tensor([[ 0.0000e+00, 0.0000e+00, 0.0000e+00, ..., 2.3447e-06,
         -1.9127e-04, -5.3330e-05]])
>>> samples.sample_rate
16000
```

1

```
ds.cast_column("audio",  
Audio(sampling_rate=16_000))
```

48 khz



2

```
ds[0]["audio"]
```

16 khz



ket hợp chung trong một bộ xử lý:

```
>>> from transformers import AutoTokenizer, AutoFeatureExtractor, AutoProcessor
```

```
>>> model_checkpoint = "facebook/wav2vec2-large-xlsr-53"
```

```
# sau khi xác định tệp vocab.json, bạn có thể khởi tạo đối tượng mã thông báo:
```

```
>>> tokenizer = AutoTokenizer("./vocab.json", unk_token="[UNK]", pad_token="[PAD]", word_delim
```

```
>>> feature_extractor = AutoFeatureExtractor.from_pretrained(model_checkpoint)
```

```
>>> processor = AutoProcessor.from_pretrained(feature_extractor=feature_extractor, tokenizer=t
```

- Đối với các mô hình nhận dạng giọng nói được tinh chỉnh, bạn chỉ cần tải bộ xử lý :

```
>>> from transformers import AutoProcessor
```

```
>>> processor = AutoProcessor.from_pretrained("facebook/wav2vec2-base-960h")
```

When you use `map()` with your preprocessing function, include the `audio` column to ensure thực ra bạn đang lấy mẫu lại dữ liệu âm thanh:

```
>>> def prepare_dataset(batch):
```

```
...audio = batch["audio"]
```

```
...batch["input_values"] = processor(audio.get_all_samples().data, sampling_rate=audio["sampl
```

```
...batch["input_length"] = len(batch["input_values"])
```

```
...with processor.as_target_processor():
```

```
...batch["labels"] = processor(batch["sentence"]).input_ids
```

```
...lô hàng trả lại
```

```
>>> dataset = dataset.map(prepare_dataset, remove_columns=dataset.column_names)
```